Intron Retention Controls Localization of lncRNAs *PURPL* and *MALAT1* to Promote Cell Proliferation and Migration

Ioannis Grammatikakis[1], Chosita Norkaew[1], You Jin Song[2], Amit K. Behera[3], Erica C. Pehrsson[4,5], Corrine Corrina R. Hartford[1], Shreya Kordale[3], Rishabh Prasanth[2], Yongmei Zhao[6], Biraj Shrethsa[6], Xiao Ling Li[1], Ravi Kumar[1], Ragini Singh[1], Tayvia Brownmiller[7], Xinyu Wen[4,5], Natasha J. Caplen[7], Pablo Perez-Pinera[8], Kannanganattu V. Prasanth[2*], Thomas Gonatopoulos-Pournatzis[3*], Ashish Lal[1*]

[1]Regulatory RNAs and Cancer Section, Genetics Branch, Center for Cancer Research (CCR), National Cancer Institute (NCI), National Institutes of Health (NIH), Bethesda, Maryland, USA

[2]Department of Cell and Developmental Biology, Cancer Center at Illinois, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA

[3]Functional Transcriptomics Section, RNA Biology Laboratory, NCI, NIH, Frederick, Maryland, USA

[4]OMICS Technology Facility, Bioinformatics, Genetics Branch, CCR, NCI, NIH, Bethesda, Maryland, USA

[5]Advanced Biomedical Computational Science, Frederick National Laboratory for Cancer Research, Frederick, Maryland, USA

[6]Sequencing Facility Bioinformatics Group, Bioinformatics and Computational Science Directorate, Frederick National Laboratory for Cancer Research, NCI, NIH, Frederick, Maryland, USA

[7]Functional Genetics Section, Genetics Branch, CCR, NCI, NIH, Bethesda, Maryland, USA

[8]Department of Bioengineering, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA

*Corresponding authors: ashish.lal@nih.gov; kumarp@illinois.edu; thomas.gonatopoulos@nih.gov

Classification: Biological Sciences

Keywords: U2AF2, lncRNA, *PURPL*, *MALAT1*, intron retention, splicing, CRISPR/Cas9, nuclear speckles, proliferation, cell migration

Abstract

Intron retention (IR) is increasingly recognized as a feature of long noncoding RNAs (lncRNAs), yet the mechanisms that shape IR in lncRNAs and the functional consequences of this process remain largely unexplored. To investigate how IR contributes to lncRNA regulation, we performed a genome-wide screen to identify factors controlling IR in the lncRNA *PURPL*. This approach uncovered a prominent role for U2AF2, which promotes retention of a specific intron in *PURPL* through a weak polypyrimidine tract. IR of this intron drives nuclear enrichment of *PURPL* and enhances cell proliferation, revealing biological relevance. Transcriptome-wide analyses showed that although U2AF2 broadly supports canonical splicing consistent with its well-established function in promoting splicing, it also facilitates IR within a distinct subset of RNAs, including the nuclear speckle–associated lncRNA *MALAT1*. Loss of U2AF2 disrupts *MALAT1* speckle localization and using *MALAT1* knockout cells reconstituted with wild-type or intron deleted variants, we identified a single intron critical for *MALAT1*'s speckle localization. Deletion of this intron from endogenous *MALAT1* impaired speckle localization and reduced cell migration, phenocopying the loss of *MALAT1*. Together, these findings reveal IR as a key regulatory mechanism governing lncRNA localization and function and uncover an unexpected role for U2AF2 in promoting IR within specific lncRNA contexts.

Introduction

Long non-coding RNAs (lncRNAs) are RNA molecules >200 nucleotides (nt) long that lack coding potential. Similar to protein-coding genes, lncRNAs are transcribed by RNA Polymerase II, are 5′-capped, spliced and are often polyadenylated (1). Several characteristics distinguish lncRNAs from protein-coding genes, including higher tissue-specific expression, lower abundance and stability, predominant nuclear localization, and having fewer but longer exons (2-6). One of the most distinguished features of lncRNAs is that they are less efficiently spliced compared to mRNAs (4,7-9). Interestingly, lncRNAs contain introns with much slower intron-excision kinetics compared to protein-coding genes (10). However, the factors that regulate lncRNA splicing and their impact on lncRNA biology remain largely unclear. We previously reported that the lncRNA *PURPL* exhibits prominent splicing defects, including frequent retention of its second intron. Transcriptomic analyses using both long-read and short-read sequencing revealed multiple *PURPL* isoforms, several of which retain intron 2 (11). While *PURPL* is also known to be transcriptionally induced by p53 and to contribute to cancer cell growth through multiple mechanisms, the regulatory factors governing intron 2 splicing and the functional significance of its retention have not been defined. This gap provides an opportunity to uncover broader principles of IR regulation in lncRNAs.

Splicing is a highly regulated process where *cis*- and *trans*-acting factors coordinate proper and efficient spliceosomal recruitment (12). Several consensus sequences in the RNA are crucial for splicing such as the polypyrimidine (Py)-tract, a stretch of pyrimidines close to the 3′ splice site (ss) (13). The splicing factor U2AF1 (U2-auxiliary factor 1) binds to the 3′ ss and U2AF2 binds to the Py-tract. The U2AF heterodimer is recruited to the branch point, and this complex serves as a critical intermediate for the subsequent assembly of the spliceosome and splicing

93 catalysis (14-18). The vast majority of human genes encode pre-mRNAs that contain introns that
94 are spliced out and most nascent transcripts undergo alternative splicing, which leads to the
95 combinatorial joining of exons. Alternative splicing is a major contributor to protein isoform
96 diversity, enabling the generation of multiple transcript variants from a single gene and frequently
97 shaping measurable cellular phenotypes (19-22). One of the least studied forms of alternative
98 splicing is intron retention (IR). Lately, it has drawn more attention since it has an important role
99 in post-transcriptional gene expression regulation, and it is believed that more than half of human
100 multi-exonic genes are affected by IR (23). IR frequently leads to nuclear detention of transcripts
101 and eventually degradation which may serve as a defense mechanism of the cell against deleterious
102 products. The extra sequence added by the retained intron may change the open reading frame
103 (ORF) of protein-coding genes and lead to nonsense-mediated decay (NMD) (24-26) or initiate
104 the use of an upstream ORF (27,28). In addition, they can contain binding sites for microRNAs
105 (miRNAs) or RNA-binding proteins (RBPs), or they may form structural changes that impact the
106 function of the transcript (29). Interestingly, IR has been associated with several diseases such as
107 cancer where tumor cells show higher levels of IR compared to normal tissues (30). Studies on IR
108 have shown that retained introns have been associated with weaker (non-consensus) splice sites,
109 higher GC content, various lengths of introns, and higher intronic sequence conservation (23,31).
110 lncRNAs have been shown to participate in splicing regulation, with *MALAT1* as the most
111 prominent example. *MALAT1* is a highly abundant nuclear lncRNA which has oncogenic activity
112 and is one of the first lncRNAs associated with cancer (32-34). It also has the ability to regulate
113 alternative splicing because it resides in the nuclear speckles and can interact with several SR
114 (Serine/Arginine rich) proteins to regulate their phosphorylation state and distribution between
115 transcription sites and speckles (33,35-37). Nuclear speckles are sub-nuclear compartments that

116 act as hubs in coordinating transcription, RNA processing and mRNA export (38). Several RBPs
117 reside in nuclear speckles such as the protein SON which is a key component and marker of nuclear
118 speckles (39). However, *MALAT1* itself is not essential for the structural integrity of nuclear
119 speckles (40,41). Nevertheless, it remains unclear whether *MALAT1* undergoes alternative splicing
120 and how this process impacts its localization to nuclear speckles.

121 Here, we show that the intron 2-retaining *PURPL* isoform, which localizes to the nucleus,
122 has a shorter half-life than the spliced isoform and promotes cell proliferation. To identify
123 regulatory factors controlling the splicing of this functionally relevant intron, we utilized our
124 recently developed CRISPR-based genome-wide screening strategy using a pooled guide RNA
125 library with a *PURPL* intron 2 minigene reporter containing its flanking exons. Unexpectedly, this
126 approach revealed a role of the splicing activator U2AF2 in promoting *PURPL* IR. U2AF2 was
127 further found to promote a broader intron retention program that includes two introns within
128 *MALAT1*. Notably, depletion of U2AF2 resulted in *MALAT1* exclusion from nuclear speckles.
129 *MALAT1* knockout cells exogenously expressing *MALAT1* truncation mutants lacking specific
130 introns identified one intron essential for the nuclear speckle association of the *MALAT1* transcript.
131 Breast cancer cells where this intron is deleted showed reduced cell migration demonstrating a
132 functional role of the intronic sequences. These findings provide insights into the splicing
133 regulation of *PURPL* and *MALAT1*, uncovering a previously unrecognized functional role of intron
134 retention in lncRNAs.

135 **Results**

136 **A CRISPR-based screen identifies U2AF2 as an activator of *PURPL* intron 2 retention**

137 We previously identified *PURPL* as a p53-regulated lncRNA that promotes the growth of
138 colorectal and liver cancer cells (11,42). In those studies, we reported that *PURPL* undergoes

139 alternative splicing, with the ~1.9 kb intron 2 either retained or spliced in the mature transcript
140 (11). To identify regulators of *PURPL* intron 2 retention, we applied our recently developed
141 screening platform, CRISPR-based identification of Regulators of Alternative Splicing with
142 Phenotypic Sequencing (CRASP-seq) (43). This approach utilizes a genome-wide knockout
143 lentiviral library targeting ~18,500 protein-coding genes using a dual hybrid guide (hg)RNA
144 system that combines *Streptococcus pyogenes* (*Sp*)Cas9 and optimized *Acidaminococcus sp.*
145 (op)Cas12a nucleases. (22,44-46). The library also includes ~2,400 pairs and ~500 curated genetic
146 interaction pairs, allowing combinatorial perturbations from a single construct. Each gene or gene
147 pair is targeted by four independent hgRNAs alongside ~500 intergenic and non-targeting controls,
148 totaling ~ 91,000 hgRNAs. The lentiviral library incorporates a doxycycline-inducible *PURPL* IR
149 reporter comprised of the full native intron 2 sequences (~1.9 kb) flanked by the adjacent exons 2
150 (161 bp) and 3 (163 bp) (Figure 1A).

151 We performed the CRASP-screen in two cell lines: HAP1 and RPE1. Following puromycin
152 selection for 48 h and doxycycline induction for an additional 24 h, RNA was extracted, and poly-
153 A RNA was purified using oligo(dT) beads to exclude transcripts still undergoing processing.
154 cDNA was synthesized using customized primers containing unique molecular identifiers (UMIs),
155 facilitating the removal of PCR duplicates during downstream analysis. The synthesized cDNA
156 was then utilized to construct Illumina sequencing libraries for paired-end sequencing. This
157 sequencing approach enables simultaneous quantification of transcripts that either retain or splice
158 out an intron of interest from one read, while the corresponding hgRNA sequences were captured
159 in the other read to pinpoint specific genetic perturbations. By directly linking *PURPL* splicing
160 outcomes with genetic modifications introduced by each hgRNA in the library, this platform
161 allowed the genome-wide identification of *PURPL* splicing regulators (Figure 1B).

162 We employed an analytical pipeline that prioritized candidate genes based on multiple
163 hgRNAs showing significant deviations from intergenic controls, with consistent identification
164 across two technical replicates. This analysis identified 62 genes that influence *PURPL* intron 2
165 splicing, the majority of which act as suppressors of intron 2 splicing (Figure 1C and Table S1).
166 Gene ontology analysis of biological processes revealed a strong enrichment for terms related to
167 mRNA processing and RNA splicing. Additionally, analysis of cellular components indicated that
168 the identified hits are predominantly associated with Spliceosomal Complex, including the U2-
169 type Spliceosomal Complex and the Catalytic step 2 Spliceosome (Figure 1D). These findings
170 validate the effectiveness of our screening approach.

171 Consistent with the endogenous *PURPL* transcript isoforms, the basal levels of IR of the
172 reporter were very high, with the majority of amplified transcripts corresponding to intron-
173 containing isoforms, as expected based on the intensity of the Illumina libraries before sequencing
174 (Figure S1A). Thus, the detection of genes that suppress IR might be below the detection levels of
175 this screen. Therefore, we focused on the genes that promote intron retention. By calculating the
176 change in Percent Intron Retention ($\Delta PIR = PIR_{guide} - PIR_{intergenic}$) across the two cell lines, we
177 identified 5 unique genes as regulators of *PURPL* intron 2 retention, all of which act to suppress
178 intron 2 splicing (Figure 1E and Table S1). This highly significant overlap between the two cell
179 lines (Figure 1E; odds ratio = 204.4, $p=2.81xe^{-10}$; Fisher's exact test) indicated a shared regulatory
180 mechanism governing *PURPL* IR, although cell-type-specific effects may also contribute.

181 Surprisingly, the gene with the strongest impact based on the average ΔPIR metric and
182 identified as a hit in both cell lines, was U2AF2 (Figure 1F). This was followed by Small Nuclear
183 Ribonucleoprotein U1 Subunit 70 (SNRNP70), a core component of the U1 snRNP, and
184 Elongation Factor Tu GTP Binding Domain, which is associated with U5 snRNP (Figure 1F). Our

185 finding that U2AF2 promotes IR of *PURPL* intron 2 was unexpected because U2AF2 is a
186 component of the spliceosome machinery, and its canonical function is to promote intron removal
187 (14,18).

188 **U2AF2 binds to the *PURPL* intron 2 transcript and promotes intron retention**

189 U2AF2, as a heterodimer with U2AF1, binds to the Py-tract and the 3′ ss, respectively, to promote
190 splicing (14,17). To elucidate its role in *PURPL* splicing, we investigated whether U2AF2 binds
191 to intron 2. Enhanced Crosslinking and Immunoprecipitation sequencing (eCLIP-seq) (47), from
192 the ENCODE project (48,49) from HepG2 cells revealed multiple U2AF2 and U2AF1 binding
193 sites not only proximal to the 3′ ss of *PURPL* intron 2 but also broadly distributed across intron 2
194 (Figure 2A). Compared to eCLIP data for other RBPs, U2AF1 and U2AF2 showed the highest
195 density of binding sites and the most prominent peaks. Specific binding of U2AF2 to intron 2 was
196 confirmed by performing RNA immunoprecipitation (RNA-IP) using a U2AF2-specific antibody
197 and primer pairs designed to detect *PURPL* transcripts. U2AF2 was found to interact with
198 transcripts containing *PURPL* intron 2, but not with the spliced *PURPL* transcripts (Figure 2B and
199 S1B).

200 To functionally validate the role of U2AF2 in *PURPL* intron 2 retention, we performed
201 RT-PCR assays from HAP1 and RPE1 cells transduced with *PURPL* minigene reporter constructs
202 and control or *U2AF2*-targeting sgRNAs. In both cell lines, we observed a striking decrease in
203 intron 2 retention upon *U2AF2* depletion as compared to control cells expressing sgRNAs targeting
204 intergenic regions (Figure 2C). As an alternative approach, we knocked down *U2AF2* using two
205 independent siRNAs and conducted semi-quantitative RT-PCR assays from SKHEP1, HepG2, and
206 normal human diploid fibroblasts W138 cells, using a primer triplet (Figures 2D-2F; Figures S2A
207 and S2B). In all tested cell lines, U2AF2 depletion increased *PURPL* intron 2 splicing, from ~20%

208 to ~40-60%, more than doubling the abundance of the spliced isoform. Given that U2AF2
209 functions as a heterodimer with U2AF1, we also knocked down *U2AF1* in SKHEP1 cells and
210 observed an effect similar to *U2AF2* knockdown on *PURPL* intron 2 retention regulation (Figure
211 S2C). We expected that U2AF1 would also come as one of the top hits in the screen but *U2AF1*
212 sgRNAs were not included in the library due to high off-target scores of the guides. These results
213 demonstrate that the U2AF complex promotes *PURPL* intron 2 retention and that both U2AF1 and
214 U2AF2 promote intron retention in *PURPL*.

215 **The intron-retained *PURPL* transcript is unstable and localizes to the nucleus**

216 We previously observed an enrichment of the intron 2-retaining *PURPL* transcript in the nucleus
217 in liver cancer cells (11). To validate these findings, we conducted single molecule RNA-FISH in
218 HCT116 cells using probes hybridizing to *PURPL* intron 2 (Figure 3A). Because *PURPL* is
219 upregulated in response to DNA damage (42), we treated the cells with hydroxyurea for 24 h to
220 induce *PURPL* expression. RNA-FISH revealed that the *PURPL* intron 2 containing transcript was
221 exclusively localized to the nucleus. To assay for subnuclear compartmentalization, we also
222 performed RNA-FISH to detect *MALAT1*, which resides in nuclear speckles (36). Intron 2-retained
223 *PURPL* transcripts did not colocalize with *MALAT1* indicating that *PURPL* is excluded from
224 nuclear speckles (Figure 3A). We further validated this result by conducting nuclear/cytoplasmic
225 fractionation assays followed by RT-qPCR and found that *PURPL* intron 2-containing transcripts
226 were almost exclusively nuclear (Figure S3A).

227 Intron retention can affect transcript stability, with longer RNA species often being more
228 vulnerable to degradation. To test if *PURPL* intron 2 effects transcript turnover, we conducted
229 RNA stability assays using the transcription inhibitor Actinomycin D (ActD) and observed that
230 the intron 2-retained *PURPL* transcript was less stable and had a significantly lower half-life than

231 the *PURPL* transcript in which intron 2 was spliced out (Figure 3B). To determine the functional
232 impact of *PURPL* intron 2 retention, we used CRISPRi to generate *PURPL*-depleted SKHEP1
233 cells by targeting 3 independent sgRNAs to the p53-response element of the *PURPL* promoter that
234 we had used in our previous study (11). We observed >80% reduction in the levels of endogenous
235 *PURPL* in these experiments (Figure 3C). We next stably transduced these *PURPL*-depleted cells
236 with a lentivirus to express exogenous *PURPL* transcript specifically containing intron 2 under the
237 control of a doxycycline (doxy)-inducible promoter and observed ~2-3-fold induction of the
238 exogenous *PURPL* intron 2 containing transcript (Figure 3D). Because we and others have
239 previously shown that *PURPL* is overexpressed in some cancers and *PURPL* promotes
240 proliferation (11,42,50,51), we next performed cell proliferation assays from the doxy-inducible
241 cells. Our results showed that cells expressing intron 2-contaning *PURPL* proliferate faster (~2-
242 fold) as compared to the no-doxy cells after 6 days of doxy treatment (Figure 2E). Induction of
243 intron 2-contaning *PURPL* transcripts did not have any effect on live cell percentage which
244 remained at ~100% (Figure S3B). Importantly, as a negative control, there was no effect of doxy
245 treatment on the proliferation of parental SKHEP1 cells (Figure S3C). To confirm that
246 proliferation increase induced by exogenous *PURPL* transcript containing intron 2 was not due to
247 the spliced isoform, we performed semi-quantitative RT-PCR. As shown in Figure 3F, doxy
248 treatment resulted exclusively in the detection of the intron-retained *PURPL* transcript. These data
249 suggest that intron 2 retention modulates *PURPL* function by sequestering the transcript in the
250 nucleus and reducing its stability, while simultaneously promoting cell proliferation, highlighting
251 a critical regulatory role for intron 2.

252 **A weak polypyrimidine tract is critical for *PURPL* intron 2 retention**

253 U2AF2 binds to the Py-tract to initiate spliceosomal assembly leading to splicing and intron
254 removal (14). Suboptimal splicing efficiency may be the result of multiple factors, including splice
255 site strength and the Py-tract sequence composition. The Py-tract is a stretch of 15-20 pyrimidines
256 (mostly Us) immediately upstream of the 3′ ss. We analyzed the Py-tract of *PURPL* intron 2 and
257 noticed that it deviates from the consensus, containing numerous purines (Figure S4A). To
258 investigate whether U2AF2 regulation on intron 2 retention relies on the Py-tract sequence, we
259 used the *PURPL* intron 2 minigene reporter system (Figure 1A). We mutated the Py-tract,
260 replacing it either with a strong Py-tract from *PURPL* intron 1 or with a stretch of Us, simulating
261 a strong Py-tract. *U2AF2* was knocked down in SKHEP1 cells stably expressing the mutated Py-
262 tract constructs and RT-PCR was conducted to monitor intron 2 splicing. Interestingly, the basal
263 levels of intron retention were very different with different mutants; ~90% in intron 2 WT (wild-
264 type), ~30% in *PURPL* intron 1 and ~20% in U-rich construct (Figure S4B and S4C). This data
265 demonstrates that the weak Py-tract is responsible for the inefficient intron 2 removal in *PURPL*
266 (Figure S4C). *U2AF2* knockdown resulted in modest increase in splicing of the mutant constructs
267 as well. Because the minigene reporter contained the mutant Py-tract along with the remaining
268 endogenous sequences from *PURPL* intron 2, the modest increase in splicing observed for the
269 mutants after *U2AF2* knockdown may reflect U2AF2 inhibitory binding at additional sites within
270 *PURPL* intron 2 beyond the Py-tract (Figure 2A). These data indicate that the endogenous intron
271 2 Py-tract in *PURPL* is a weak Py-tract that controls the high intron retention observed in *PURPL*.

272 **U2AF2 directly regulates multiple IR events including *MALAT1***

273 To determine if U2AF2 promotes IR in additional genes, we next performed total RNA-seq and
274 PacBio Iso-seq from SKHEP1 cells transfected with siCTRL or siU2AF2. Consistent with our
275 data, *U2AF2* knockdown resulted in a marked increase in the splicing efficiency of *PURPL* intron

276 2 (Figures S5A and S5B). Additionally, *U2AF2* depletion increased overall *PURPL* expression,
277 with increased read coverage across all exons (Figures S5A and S5B; Tables S3 and S4). To
278 identify IR events regulated by U2AF2 at a transcriptome-wide level, we utilized IRFinder, an
279 algorithm specifically designed to calculate changes in IR (52). This analysis revealed ~1000
280 statistically significant IR events regulated by U2AF2. For majority of these events (732), there
281 was an increase in the IR ratio (Intronic abundance/(Intronic abundance + exonic abundance)) after
282 *U2AF2* depletion, indicating that U2AF2 predominantly promotes splicing, as expected. However,
283 in 269 events, *U2AF2* knockdown resulted in decreased IR, indicating that U2AF2 promotes IR in
284 a subset of genes (Figure 4A and Table S2). We next analyzed ENCODE U2AF2 eCLIP-seq data
285 from HepG2 cells to determine if U2AF2 directly binds to the regulated transcripts. U2AF2
286 binding was detected in 96 out of the 269 IR events supporting a direct role for U2AF2 in
287 promoting IR (Figure S5C and Table S2).

288 To validate these findings, we selected three of the identified introns from *MDM1*,
289 *RETREG2*, and *TMEM41B* (Figure S6A). We observed multiple U2AF2 binding peaks within the
290 regulated introns from *MDM1* (Figure S6B). We next assessed their splicing in SKHEP1 cells
291 following *U2AF2* knockdown using two independent siRNAs. In all three cases, *U2AF2* depletion
292 significantly enhanced intron removal (Figures S6C-S6E). IRFinder validated that *PURPL* intron
293 2 was one of the events showing a significant decrease (~2-fold) in IR ratio upon *U2AF2* depletion
294 (Figure 4B).

295 Among the lncRNAs in which IR is promoted by U2AF2, was the lncRNA *MALAT1*.
296 IRFinder identified two introns in *MALAT1* that have weak Py-tract sequences similar to *PURPL*
297 intron 2 (Figure S7A-S7C). For both introns, *U2AF2* knockdown resulted in a significant decrease

298 (~2-fold) in the IR ratio compared to siCTRL (Figures 4A and 4B). This result was confirmed by
299 RT-PCR from HCT116 and SKHEP1 cells upon *U2AF2* knockdown (Figures 4C-4F).

300 **Intron 2 of *MALAT1* drives its localization to nuclear speckles**

301 *MALAT1* regulates splicing by interacting with splicing factors in nuclear speckles.(36) To
302 determine the impact of IR on *MALAT1* localization to nuclear speckles, we performed RNA-FISH
303 for *MALAT1* upon *U2AF2* depletion in SKHEP1 and HCT116 cells. Remarkably, *U2AF2*
304 knockdown resulted in a significant decrease in the ratio of *MALAT1* signal intensity in speckles
305 relative to nucleoplasm (Figures 5A and 5B; Figures S8A and S8B). U2AF2 depletion did not
306 affect total *MALAT1* levels as observed in our RNA-seq (Table S3). This reduction in *MALAT1*
307 levels in nuclear speckles did not compromise the structural integrity of nuclear speckles as
308 indicated by immunostaining for the RNA-binding protein SON, a well-established nuclear
309 speckle marker (39) (Figures 5A and S8A) and is consistent with previous reports showing that
310 *MALAT1* is not essential for the integrity of nuclear speckles (40,41). These data suggest that
311 U2AF2 plays an important role in promoting *MALAT1* localization to nuclear speckles.

312 To determine whether it is the intron retention in the transcript *per se* which directs
313 *MALAT1* localization or if U2AF2 indirectly regulates its localization, we assessed the nuclear
314 speckle localization of wild-type (WT) and intron-deleted *MALAT1* transcripts. To do this, we
315 used CRISPR/Cas9 to generate two *MALAT1* knockout (KO) clones in HCT116 by inserting ~100
316 bp downstream of the transcription start site, a puromycin resistance gene in one allele and a
317 hygromycin resistance gene in the other allele. In both KO clones, >95% reduction in *MALAT1*
318 RNA levels was observed as compared to a wild-type clone (Figure S9A).

319 Next, we transduced the *MALAT1* KO clones with an empty vector or doxycycline-
320 inducible lentiviral constructs that exogenously express full-length *MALAT1*-WT in which both

321 introns are intact, intron 1 deleted *MALAT1* (del-1), intron 2 deleted *MALAT1* (del-2) or both
322 introns were deleted (del-1+2). After 48 h of doxy-treatment to induce *MALAT1* expression, which
323 was confirmed with RT-qPCR (Figure S9B), RNA-FISH combined with immunofluorescence for
324 the nuclear speckle marker SON, revealed distinct localization patterns. As expected, exogenously
325 expressed *MALAT1*-WT predominantly (~80%) localized to nuclear speckles (Figures 6A and
326 S10A). Similarly, *MALAT1*-del-1 also predominantly (~80%) localized to nuclear speckles. In
327 contrast, *MALAT1* lacking intron 2 (del-2) or both introns (del-1+2) exhibited a marked reduction
328 (<15%) in nuclear speckle localization and instead displayed diffused localization in the nucleus
329 (Figures 6A and S10A). Importantly, this diffused localization upon intron 2 deletion was observed
330 in both *MALAT1* knockout clones and was further confirmed through quantitative analyses
331 (Figures 6B and S10B).

332 **Deletion of *MALAT1* Intron 2 using CRISPR/Cas9 impairs cell migration**

333 *MALAT1* has been shown to promote cell migration and metastasis (32,53) and its expression has
334 been correlated with high tumor progression and metastasis in various tumors (54). We therefore
335 tested whether the intronic sequences that regulate *MALAT1* nuclear speckle localization are
336 related to the ability of *MALAT1* to promote cell migration. To investigate this, we used the
337 metastatic breast cancer cell line MDA-MB-231 and first performed RNA-FISH for *MALAT1* after
338 knocking down *U2AF2*. Similar to SKHEP1 and HCT116 cells, *U2AF2* knockdown resulted in
339 decreased localization of *MALAT1* to speckles in MDA-MB-231 cells (Figures 7A and 7B). We
340 next utilized CRISPR/Cas9 to generate KO clones in MDA-MB-231 cells where either the whole
341 *MALAT1* locus or intron 2 of *MALAT1* were deleted. *MALAT1* depletion in the *MALAT1* KO clone
342 was confirmed with RT-qPCR (Figure S11A). RNA-FISH in 2 separate clones showed that intron
343 2 deletion reduced *MALAT1* localization to nuclear speckles compared to the WT clone (Figures

344 7C-7E). In the intron-2 deletion clones, the transcripts containing intron 2 were, as expected,
345 undetected with primers specifically amplifying intron 2. Interestingly, there was a modest
346 increase in total *MALAT1* expression suggesting a compensatory mechanism for the loss of intron
347 2 (Figure S11B). Functionally, when we tested the migration potential of these cells in transwell
348 cell migration assays, we found significantly reduced cell migration in the intron 2-deleted clones,
349 phenocopying the complete loss of *MALAT1* locus (Figures 7F and S11C). These data suggest that
350 the retention of intron 2 is essential for *MALAT1* localization to nuclear speckles and it plays a key
351 role in promoting cell migration. Together, our data uncovers a U2AF2-driven intron retention
352 program in lncRNAs that regulates cell proliferation and migration.

353 **Discussion**

354 The goal of this study was to examine the causes and consequences of IR in lncRNAs, focusing
355 on *PURPL* and *MALAT1*. Intron retention is a regulated process that shapes gene expression
356 networks through mechanisms such as NMD and influences alternative protein isoform generation
357 during differentiation, development, and disease (25,55). We previously identified a p53-induced
358 lncRNA, *PURPL*, which exhibits growth-promoting functions, and that intron 2 of *PURPL* is
359 retained in multiple cell lines (11). To accomplish our goal, we employed an innovative RNA-seq
360 coupled genome-wide CRISPR screen on a reporter containing *PURPL* intron 2 and unexpectedly
361 identified U2AF2 as a major inhibitor of intron 2 removal in *PURPL*. U2AF2 forms a heterodimer
362 with U2AF1, binding to the Py-tract and the 3′ ss, respectively, to promote splicing by recruiting
363 core spliceosomal components (17,18,56). Therefore, given the established role of U2AF2 as a
364 critical positive regulator of splicing catalysis, we were surprised that U2AF2 inhibits the removal
365 of *PURPL* intron 2. While U2AF1 was not identified as a hit in our screen, follow-up experiments

366 revealed that U2AF1 also suppresses *PURPL* intron 2 removal, suggesting a previously
367 unrecognized role for the U2AF complex as a negative regulator *PURPL* IR.

368 Beyond IR in *PURPL*, we sought to determine whether U2AF2 promotes IR in other
369 transcripts. Using IR Finder (52), we identified other genes where U2AF2 promoted IR. One of
370 the genes that drew our attention was the lncRNA *MALAT1* which showed two intron retention
371 events regulated and bound by U2AF2. These introns contain weak splice sites suggesting that this
372 may be the reason that they are regulated, and they are not constitutively spliced. Because *MALAT1*
373 interacts with splicing factors within nuclear speckles (36), we examined whether U2AF2 affects
374 *MALAT1* localization to speckles. Indeed, our results establish *MALAT1* intron 2 as a critical
375 determinant of its nuclear speckle localization. Future studies involving mapping of the intronic
376 sequences and targeted mutagenesis of the splice sites within *MALAT1* intron 2 could help
377 distinguish whether speckle localization of *MALAT1* is inhibited by the splicing process itself or
378 depends on a discrete element within the intron.

379 For *PURPL*, previous work established its context-dependent oncogenic activity (11,42),
380 but those studies depleted all transcripts from the *PURPL* locus despite the presence of multiple
381 alternatively spliced isoforms. In the current study, overexpression of the intron 2-retaining
382 *PURPL* isoform in CRISPRi-mediated *PURPL*-depleted cells increased proliferation,
383 demonstrating isoform-specific, growth-promoting function. Our findings reveal that intron
384 retention in both *MALAT1* and *PURPL* is a regulatory mechanism with direct consequences on
385 lncRNA localization and cancer-related phenotypes.

386 The surprising finding of U2AF2 promoted IR promoted us to determine the underlying
387 mechanism. We first asked whether U2AF2 influences transcription initiation or elongation at the
388 *PURPL* locus, given the tight coupling between transcription and splicing (57,58). Transcription

389 kinetics can modulate IR (23,59) and lncRNAs often lack proximal RNA-pol II phosphorylation
390 over 5′ ss (10). We therefore assayed for PolII-pSer5 (transcription initiation) and PolII-pSer2
391 (transcription elongation) occupancy after U2AF2 knockdown using CUT&RUN-Seq but there
392 was no significant difference around intron 2 or elsewhere across the *PURPL* locus (data not
393 shown). We also examined the chromatin mark H3K36me3 which has been linked to IR (60) and
394 again observed no effect after U2AF2 depletion (data not shown).

395 U2AF2 functions by binding to the Py-tract, and the Py-tract of *PURPL* intron 2 and
396 *MALAT1* introns 1 and 2 deviate from the consensus, containing a mix of pyrimidines and purines.
397 These atypical sequences likely contribute to splicing inhibition as revealed by our data from the
398 mini-gene reporter for *PURPL*. Given that U2AF1 and U2AF2 bind to multiple intronic sites, a
399 plausible model is that their docking obstructs subsequent spliceosomal assembly. lncRNAs
400 generally undergo less efficient splicing and consequently, they have higher levels of IR compared
401 to protein coding genes (8,10). A study showed that lncRNA splicing strength correlates with 5′ ss
402 strength and lncRNA Py-tract has higher thymidine content compared to protein coding genes
403 (61). The reduced evolutionary pressure to maintain efficient splicing as they are not used as
404 templates for protein synthesis, likely contributes to this pattern. This is also supported by their
405 predominant nuclear localization (3,4).

406 Our data shows that *PURPL* transcripts can be either nuclear or cytoplasmic, and U2AF2
407 regulation of IR offers a mechanism of nuclear detention. Similar to this notion, the lncRNA *TUG1*
408 remains in the nucleus due to intron retention (62), and the mouse lncRNA *pCharme* undergoes
409 an IR event which renders it chromatin-bound through its association with MATR3 and PTBP1
410 (63). Another aspect of the effects of IR lies in the turnover of the transcripts where additional
411 sequences added by the introns may cause lower transcript stability. In support of this, nuclear

412 lncRNAs have been shown to be less stable than cytoplasmic ones (5), and *PURPL* also follows
413 this rule. Notably, the human lncRNA *hFAST* is cytoplasmic and promotes pluripotency through
414 WNT signaling, but the mouse homolog does not because of its nuclear localization (64).

415 Introns have been proposed to act as docking elements that tether lncRNAs to chromatin
416 (4,63), and a recent study showed that nuclear speckles are associated with distinct classes of
417 retained introns (65). This raises the question of what makes intron 2 important for speckle
418 localization of *MALAT1*. One possibility is that this region recruits specific RBPs or adopts a
419 structural configuration required for speckle association. Our data further indicates that intron 2
420 retention contributes to *MALAT1's* role in cell migration, potentially because speckle localization
421 is necessary to sustain this phenotype. Future studies are warranted to dig deeper into the
422 underlying mechanism and to determine how intron retention shapes the function of *PURPL* and
423 *MALAT1*.

424 Collectively, our findings reveal an expected function of U2AF2 in promoting intron
425 retention that governs the subcellular localization and function of the lncRNAs *PURPL* and
426 *MALAT1.* As intron retention continues to emerge as a regulated form of alternative splicing, the
427 evidence presented here expands the framework for the mechanisms and biological consequences
428 of intron retention.

429 **Materials and Methods**

430 **Cell culture and treatment**

431 HAP1, HCT116, HEK293T, HepG2, RPE1, SKHEP1, U2OS, WI38, and MDA-MB-231 cells
432 were purchased from ATCC and maintained in Dulbecco's Modified Eagle Medium (DMEM)
433 (Gibco) supplemented with 10% (v/v) fetal bovine serum (Gibco) and 1% (v/v) PenStrep (Gibco)

434 in a 5% CO_2 atmosphere at 37°C. All cell lines were routinely confirmed to be free of mycoplasma
435 using Venor GeM Mycoplasma Detection Kit (Millipore Sigma-Aldrich).

436 **CRASP-Seq *PURPL* screen**

437 The *PURPL* minigene reporter, composed of exons 2 and 3 with the full native intronic sequence,
438 was cloned into the CRASP-Seq genome-wide library as described previously (43). Reporter
439 fragments were synthesized by TWIST Biosciences and assembled into the Eco32I-digested
440 CRASP-Seq knockout library using NEBuilder (46). Library coverage (>250-fold) was maintained
441 through parallel cloning reactions and electroporation into Endura competent cells followed by
442 plasmid amplification and extraction.

443 Lentivirus was produced following established protocols (46). The pooled library was used
444 to infect HAP1 and RPE1 cells expressing *Sp*Cas9 and opCas12a nucleases (46). at a multiplicity
445 of infection (MOI) of ~0.2. After puromycin selection (2 µg/mL for 48 hours), cells were pooled,
446 induced with doxycycline to activate reporter expression and harvested 24 hours later for RNA
447 analysis.

448 Total RNA was extracted, poly(A)+ RNA isolated, and cDNA synthesized using a U6-
449 annealing primer. Spliced and intron-retaining reporter transcripts were amplified using exon- or
450 intron-specific primers incorporating Illumina barcodes. Libraries were sequenced on the NovaSeq
451 6000 platform with a 25% PhiX spike-in and a 300-cycle kit. The sequencing strategy consisted
452 of the following configuration: Read 1 (210 cycles), Index Read 1 (8 cycles), Index Read 2 (8
453 cycles) and Read 2 (85 cycles). Each sample yielded over 500 million total reads.

454 **CRASP-seq data analysis**

455 CRASP-Seq data were processed using a custom pipeline (43). Guides, cell barcodes, and UMIs
456 were retrieved using STAR, and alignment were managed with SAMTools. Reads containing

457 complete hybrid guide, barcodes, UMI, and splicing outcomes were deduplicated prior to analysis.
458 Splicing outcomes were quantified as percent intron retention (PIR): PIR = [IE / (IE + EE)] × 100,
459 where IE represents intron retention reads, and EE represents exon-exon junction reads.
460 Differential PIR (ΔPIR) was calculated as: $\Delta PIR = PIR_{guide} - PIR_{intergenic}$

461 The ΔPIR for each gene and for each guide separately is shown in Table S1. Guides with
462 fewer than 10 reads were excluded. A ΔPIR threshold corresponding to 10% FDR (based on
463 intergenic controls) defined significant hits. At the gene level, hits required FPKM ≥ 0.1, at least
464 two qualifying guides, and ≥50% of guides exceeding the ΔPIR threshold in both technical
465 replicates. Genes identified through combinatorial targeting that met these criteria were included
466 as hits (Table S1). The heatmap and Venn diagram in Figure 1C and 1E display single-gene hits
467 as well as collapsed gene hits resulting from combinatorial targeting.

468 **eCLIP-seq data and functional enrichment analysis**

469 eCLIP-seq data for U2AF2, U2AF1, PTBP1, PRPF8, and SRSF1 were obtained from the
470 ENCODE database (encodeproject.org) with the following identifiers: ENCFF989JBA (CTRL),
471 ENCFF566CFJ (CTRL), ENCFF536AFD (U2AF2), ENCFF913WRH (U2AF2), ENCFF159SPZ
472 (U2AF2), ENCFF368XEI (U2AF2), ENCFF542NOA (U2AF1), ENCFF651VRQ (U2AF1),
473 ENCFF363UDO (PTBP1), ENCFF044RKN (PRPF8), and ENCFF327NVE (SRSF1). For
474 visualization of RNA-seq data, the Integrated Genome Viewer (IGV) version 17 was used. The
475 functional enrichment analysis was performed using g:Profiler (66) and the results are shown in
476 the 'gProfiler_Results' tab in Table S1.

477 **siRNA transfections**

478 For siRNA transfections, we used AllStars Negative Control siRNA (QIAGEN). U2AF2 siRNAs
479 were purchased from Dharmacon (J-012380-18 and J-012380-19). U2AF1 siRNA was also

480 purchased from Dharmacon (J-012325-10). Cells were reverse transfected with siRNA at a final
481 concentration of 20 nM using Lipofectamine RNAiMAX (Invitrogen) in Opti-MEM I Reduced
482 Serum Medium (Gibco) according to the manufacturer's protocol. SKHEP1 cells were transfected
483 for 72 hr before RNA harvesting. U2OS, HepG2 and WI38 cells were transfected for two rounds
484 of 48 h + 72 h or 72 h + 72 h. For RNA FISH, SKHEP1, HCT116, and MDA-MB-231 cells were
485 transfected for 72 hr. SKHEP1 expressing the *PURPL* Py-tract mutants were transfected for two
486 rounds of 48 h and 24 h before harvesting, the cells were treated with 1 μg/mL doxycycline to
487 induce the transgene expression. For RNA-seq, the cells were transfected for 72 h before harvested.
488 For RNA stability assay, the cells were reseeded in 12-well plates 48 h after siRNA transfection
489 and 24 h later they were treated with Act D for the indicated time points at 5 μg/mL before RNA
490 extraction and subsequent assays. For RNA-IP, the cells were transfected for 72 h before harvested
491 in 10 cm-plates.

492 **RNA extraction, RT-qPCR, semi-quantitative RT-PCR, and splicing gels**

493 Total RNA was extracted using TRIzol (Invitrogen) according to the manufacturer's protocol.
494 cDNA was then made using iScript™ Reverse Transcription Supermix (Bio-Rad). For real-time
495 qPCR, all reactions were carried out on StepOnePlus real-time PCR System (Applied Biosystems)
496 using FastStart SYBR Green Master Mix (Millipore Sigma). *GAPDH* mRNA was used to
497 normalize expression except for RNA-IPs in Figure 2B where *18S* was used for normalization, and
498 the relative expression/enrichment of RNAs was calculated using the $2^{-\Delta\Delta Ct}$ method. RT-qPCR
499 primer sequences for each gene are indicated in Table S3. For splicing gels, cDNA was used to
500 amplify the region flanking the retained intron with primers in the flanking exons with the addition
501 of one primer inside the intron as indicated in Figure S1B. In the case of *MALAT1*, primers in the
502 flanking exons were utilized due to the short length of introns. For the assay of the effect of U2AF2

503 on the Py-tract mutants of *PURPL*, primers on the transgene were used instead of the endogenous
504 *PURPL* exon 2 and 3. The sequences of the primers can be found in Table S5.

505 **RNA-IPs**

506 For RNA-IP, two 10 cm-plates of SKHEP1 cells were lysed with RIPA buffer (Thermo Scientific)
507 for 10 min on ice before spin down for 10 min at 4ºC. The lysate was incubated with 1 µg of anti-
508 U2AF2 (CST) or rabbit IgG (CST) in 500 µL total volume overnight at 4ºC. The next day, 50 µL
509 of Protein A/G Dynabeads (Pierce) were added to the mix and incubated for 2 h at 4ºC. Beads
510 were then washed 4 times with RIPA buffer and RNA was isolated using TRIzol. Equal volumes
511 of extracted RNA were used for cDNA synthesis and RT-qPCR as described above.

512 **Immunoblotting**

513 For immunoblotting, total cell lysate was prepared using RIPA buffer (Thermo Scientific)
514 containing protease inhibitor cocktail (Roche). Protein concentration was determined using the
515 BCA protein quantitation kit (Thermo Scientific). Cell lysate was loaded onto SDS-PAGE gel and
516 transferred to a PVDF membrane (Thermo Scientific) using a semi-dry apparatus (Bio-Rad). The
517 membrane was blocked in 5% skim milk (Millipore) in TBS containing 0.05% Tween 20. The
518 following antibodies were used: anti-U2AF2 (1:1000, rabbit) (CST, 70471S), anti-GAPDH
519 (1:10,000, rabbit) (CST, 2118L) and secondary anti-rabbit (1:5000) (CST, 7074S).

520 **Nuclear cytoplasmic fractionation**

521 For nuclear cytoplasmic fractionation, SKHEP1 cells were harvested after trypsinization at 80-
522 90% confluence. Cells were pelleted and resuspended in 500 µL RSB buffer (10 mM Tris-HCl,
523 100 mM NaCl, and 2.5 mM MgCl2) containing 40 µg/mL digitonin (Invitrogen), and 100 U of
524 RNase Out. After incubation on ice for 10 min, cell viability was tested with trypan blue to make
525 sure that the cells were permeabilized. The cytoplasmic fraction was harvested as supernatant after

526 centrifugation at 1,500 × g for 5 min at 4°C. The nuclear pellet was resuspended in 100 µL RSB
527 buffer. 100 µL of the cytoplasmic fractionation were transferred to a fresh tube and 900 µL of
528 TRIzol was added as well as to the nuclear extracts for RNA extraction and subsequent cDNA
529 synthesis and RT-qPCR.

530 **Cell proliferation assay**

531 The cell proliferation assay was conducted as previously described (67). To measure proliferation,
532 parental SKHEP1 cells and *PURPL*-CRISPRi cells overexpressing *PURPL* containing intron 2
533 were seeded at 1x10^5 cells per well in 6-well plates in the presence of 1 µg/mL doxycycline. After
534 3 days, the cell numbers and viability were measured. The cells were then passaged at $1x10^5$ and
535 measured again at 6 days, and cumulative cell numbers were plotted.

536 **RNA-seq**

537 The TruSeq Stranded mRNA libraries were pooled and sequenced on one NovaSeq6000 S1
538 flowcell using 2x101 cycles for pair-end run. The Real Time Analysis software (RTA v3.4.4) was
539 used for processing raw data files, the Illumina bcl2fastq v2.20 was used to demultiplex and
540 convert binary base calls and quality scores to fastq format. The sequencing reads were trimmed
541 of adapters and low-quality bases using Cutadapt (v1.18). The trimmed reads were mapped to
542 human reference genome (hg38) and Gencode annotation GENCODE_v30 GTF using STAR
543 aligner (version 2.7.0f) with two-pass alignment option. RSEM (v1.3.1) was used for gene and
544 transcript quantification based on GENCODE annotation file. Palantir Foundry was used for
545 analysis of Differential Expression of Genes inside the secure NIH Integrated Data Analysis
546 Platform (NIDAP). The DEG Analysis results are shown in Table S3.

547 **Iso-Seq**

548 The libraries were prepared using PacBio standard Iso-seq library prep using SMRTbell prep kit
549 3.0. The libraries were sequenced on PacBio Sequel II platform using PacBio v2.0 chemistry.
550 PacBio SMRTlink raw subreads were converted into HiFi circular consensus sequences (CCS)
551 with minimum 3 passes. PacBio IsoSeq v3 pipeline was used to demultiplex the barcodes and
552 remove primers. Additional refine steps included trimming poly-A tails and removing concatemers
553 to generate Full Length Non-Concatemer (FLNC) reads. The FLNC reads were used to map to
554 human reference (hg38) using Minimap2 software to generate alignment bam files. Hierarchical
555 clustering and merging steps were performed to obtain consensus isoforms and the full-length (FL)
556 consensus sequences. The high-quality FL transcripts were mapped to the reference genome by
557 using the minimap2 software. Isoform classification and quality control were done using
558 SQANTI3 software and Illumina short-read gene expression data was combined with PacBio Iso-
559 seq for transcript quantification. The Iso-Seq Analysis results are shown in Table S4. The resulting
560 bam files from the Iso-seq analysis were used for plotting and visualization with the R package
561 Gviz (https://bioconductor.org/packages/release/bioc/html/Gviz.html).

562 **IRFinder**

563 IRFinder (52) was used to identify intron retention events that changed significantly on U2AF2
564 knockdown. A docker image was downloaded from Docker Hub (docker://cloxd/irfinder:2.0.1).
565 The reference was built using pre-downloaded resources (IRFinder BuildRefProcess), using
566 Ensembl release 92 and the 100 bp mappability file provided by IRFinder. The Build-BED-refs.sh
567 script was adjusted so that transcript biotype was not restricted to "processed_transcript" or
568 "protein_coding". Intron retention was quantified using IRFinder in FastQ mode with adapter
569 trimming. Differential intron retention events between siU2AF2 vs. siCTRL were identified with

570 IRFinder Diff using DESeq2 with the default IR ratio ≥ 0.05 in at least one sample and no warning
571 filters, which profiled 16,859 unique introns.

572 **Construct generation**

573 For Py-tract mutants, gene fragments were purchased from Twist Biosciences and cloned into the
574 pLCHKO-puro vector containing the WT sequences of *PURPL* (Figure 1A), between sites for
575 NseI and AfeI. The *PURPL* transcript containing intron 2 was cloned into the pCW57 vector
576 (Hygro) (Addgene #80922) in SalI and MluI sites. *MALAT1* transcripts were cloned into the
577 pCW57 vector (BSD) (Addgene #80921) in SalI and MluI sites. The *PURPL* insert was generated
578 by Twist Biosciences and the *MALAT1* sequence was amplified using Phusion DNA polymerase
579 (NEB). For deletion mutants, we used the Q5® Site-Directed Mutagenesis Kit (NEB). The primers
580 used for cloning are listed in Table S5.

581 **Generation of *MALAT1* KO cell lines**

582 For HCT116 *MALAT1* knockout, we used integration by non-homologous end joining, which was
583 accomplished by introducing a simultaneous double-strand break in genomic DNA and in the
584 targeting vectors (pCMV-Puro and pCMV-Hygro) (68,69). Plasmids encoding spCas9 and
585 sgRNAs were obtained from Addgene (Plasmids #41815 and #47108). Oligonucleotides for
586 construction of sgRNAs were obtained from Integrated DNA Technologies, hybridized,
587 phosphorylated and cloned into the sgRNA plasmid or targeting vector using BbsI sites (70).
588 Target sequences for sgRNAs are provided in Table S5. These vectors were transfected into
589 HCT116 using Lipofectamine 2000 following recommendations from the manufacturer. Three
590 days after transfection, the cells were selected with Puromycin (0.5 μg/mL) and Hygromycin (100
591 μg/mL) to generate clonal populations. Genomic DNA from each clone was isolated using DNEasy

592 Blood and Tissue Kit (Qiagen). PCRs to detect integration of the targeting vector at the target site
593 were performed using KAPA2G Robust PCR kits (Kapa Biosystems) according to the
594 manufacturer's instructions. A typical reaction contained 20-100 ng of genomic DNA in Buffer A
595 (5 µL), Enhancer (5 µL), dNTPs (0.5 µL), primers forward (PINCR Det FP, 1.25 µL) and reverse
596 (Targeting vector Det RP, 1.25 µL) and KAPA2G Robust DNA Polymerase (0.5 U). The DNA
597 sequences of the primers for each target are provided in Table S5. PCR products were visualized
598 in 2% agarose gels and images were captured using a ChemiDoc-It2 (UVP).

599 To generate MDA-MB-231 whole-locus and intron 2 *MALAT1* knockout clones, we used
600 the CRIPSR/Cas9 system by transfecting the cells with gRNAs and purified Cas9. 2 µL of each
601 30 µM gRNA was mixed with 1 µL (~5 µg) of purified Cas9 and 100 µL of 4D-Nucleofector® X
602 SE buffer SE from the Cell Line Kit (Lonza) supplemented with 18.5 µL of SE supplement per
603 100 µL total buffer volume. The mixture was incubated at room temperature for 10 min and
604 1x10^6 cells were resuspended in nucleofection cuvettes. Nucleofections were carried out using
605 the 4D-Nucleofector® System (Lonza). The cells were transferred to a 6-well plate, and they were
606 single-cell sorted 48 h later. The clones were grown for 3-4 weeks before genomic DNA extraction
607 and genotyping with PCR using the primers in Table S5.

608 ***PURPL* and *MALAT1* overexpression and transduction**

609 *PURPL* intron 2 constructs were transduced in previously generated *PURPL*-CRISPRi SKHEP1
610 cells (11). *MALAT1*-WT and deletion mutants were transduced in the two HCT116-*MALAT1* KO
611 clones. For lentivirus generation, HEK293T cells were seeded at 2×10^5 cells/well onto 6-well
612 plates, and 1,200 ng of each construct was transfected with lentiviral package vectors using
613 Lipofectamine 2000 (Life Technologies, Invitrogen) according to the manufacturer's protocol.
614 Virus-containing media were harvested at 48 and 72 h post-transfection and added to cells for 24

615 h. Transduced cells were selected with 500 µg/mL hygromycin for 2 weeks (SKHEP1) or 10
616 µg/mL Blasticidin (HCT116).

617 **RNA-FISH, IF and Quantification of FISH**

618 The *PURPL* intron probe set was custom-designed using the Stellaris Probe Designer, tagged with
619 Quasar 570 dye and purchased from Stellaris. The *MALAT1* smFISH probe set with Quasar 670
620 dye was purchased from Stellaris (Cat# VSMF-2211-5). For *PURPL* intron smRNA-FISH,
621 SKHEP1 cells were seeded on coverslips and cultured for 24 hours at 37°C with 5% CO_2 before
622 inducing DNA damage by incubating the cells with 2 mM hydroxyurea for 24 hours (or regular
623 growth media for control). For *MALAT1* smRNA-FISH, HCT116 and MDA-MB-231 *MALAT1*
624 KO cells were seeded on coverslips and cultured for 24 hours at 37°C with 5% CO_2. *MALAT1* WT,
625 *MALAT1* del-1, *MALAT1* del-2, *MALAT1* del1+2, or empty vector expression was induced by
626 incubating the cells with 1 µg/mL doxycycline for 48 hours. The cells that were transfected with
627 siRNA for U2AF2 were seeded on coverslips at the time of transfection. Cells were fixed with 4%
628 PFA for 15 min at room temperature, permeabilized with 0.5% Triton X-100, and washed with
629 washing buffer (10% formamide, 2XSSC) for 5 min. Probe was added to hybridization buffer
630 (10% formamide in Stellaris hybridization buffer Cat# SMF-HB1-10) at a final concentration of
631 125 nM. Hybridization was done in a humidified chamber in the dark overnight at 37°C. After
632 hybridization, the coverslips were washed twice with wash buffer for 30 min at 37°C and then
633 post-fixed with 4% PFA for 10 min at room temperature. The coverslips were blocked with 5%
634 normal goat serum and then incubated with SON antibody (Sigma Cat# HPA023535). Primary
635 and secondary antibodies were diluted in 1% normal goat serum. Antibody incubations were done
636 at room temperature for 1 h and washed with PBS. DNA was counterstained by DAPI. The
637 coverslips were then washed with 4XSSC for 5 min at room temperature and mounted in

638 VectaShield Antifade Mounting Medium (Vector Laboratories). Images were taken using
639 DeltaVision microscope (GE Healthcare) equipped with 60 X/1.42 NA oil immersion objective
640 (Olympus) and CoolSNAP-HQ2 camera. To quantify the speckle to nucleoplasm signal intensity
641 ratio of *MALAT1* in Figures 5B, 7B, 7D, and S8B ImageJ was used to create nucleus masks based
642 on DAPI and speckles mask based on SON. The signal intensity of *MALAT1* in the speckles mask
643 was divided by the *MALAT1* signal intensity in the nucleus-minus-speckles area. For the
644 quantification of Figures 6B, 7E and S10B, the cells with *MALAT1* FISH signal showing
645 prominent speckle patterns similar to SON were classified as "Speckle" pattern, and the cells with
646 a diffused *MALAT1* FISH signal were classified as "Diffused". Only *MALAT1*-positive cells were
647 counted for quantification.

648 **Transwell cell migration assays**

649 For the transwell assay, WT and KO MDA-MB-231 clones were seeded at $10x10^5$ cells per well
650 in 12-well plates. 24 h later, the media were removed and replaced with media containing 1% FBS.
651 After 5 h, the cells were trypsinized and seeded in the top chamber of the transwell plates (Corning,
652 Catalog#354578) at 25,000 cells per well at a volume of 100 µL. The lower chamber of the
653 transwell is filled with 700 µL of media containing 10% FBS. The cells were incubated for 24 h
654 and the transwell were removed from the plates. The non-migrated cells were scraped with a cotton
655 swab, and the migrated cells were fixed with ice-cold methanol for 5 min and stained with 0.05%
656 crystal violet blue. The migrated cells were counted under a light microscope.

657 **Statistics**

658 Statistics were performed using the Student's t-test.

659 **Data availability**

660 The Iso-Seq data files have been deposited at GEO with accession number GSE288684 and
661 reviewer token: ktcnasayrbgnjev (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi). The RNA-
662 Seq data files have been deposited at GEO with accession number GSE288685 and reviewer token:
663 qvwhuawixxqjnml (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi).

664 **Acknowledgments**

665 This research was supported by the Intramural Research Program (A.L.) of the National Cancer
666 Institute (NCI), Center for Cancer Research (CCR) (project numbers ZIA BC011646, ZIA
667 BC012019), NIH. KVP laboratory is supported by grants from NIH (R01-GM132458), Cancer
668 center at Illinois seed grant, ARPA-H, and National Science Foundation (NSF) center for
669 quantitative cell biology. YJS is supported by NIH T32EB019944. We thank the CCR Genomics
670 Core, CCR, NCI, Bethesda, MD for its valuable assistance with Sanger and Nanopore sequencing.
671 We also thank the Sequencing Facility, CCR, NCI, Frederick, MD for performing the RNA
672 sequencing assays. The contributions of the NIH author(s) were made as part of their official duties
673 as NIH federal employees, are in compliance with agency policy requirements, and are considered
674 Works of the United States Government. However, the findings and conclusions presented in this
675 paper are those of the author(s) and do not necessarily reflect the views of the NIH or the U.S.
676 Department of Health and Human Services.

677 **Author Contributions**

678 IG, CN, YJS, CCH, SK, RP, XLL, RK, RS, and PPZ conducted experiments. IG, AKB, EP, YZ,
679 BS, TB, XW, TGP, KP, and AL did data analysis. IG and AL wrote the manuscript. NC, KP, and
680 TGP edited the manuscript.

681 **Declaration of interests**

682 The authors declare they have no competing interest.

683 **References**

684 1. Quinn, J.J. and Chang, H.Y. (2016) Unique features of long non-coding RNA biogenesis and
685 function. *Nat Rev Genet*, **17**, 47-62.
686 2. Ulitsky, I. and Bartel, D.P. (2013) lincRNAs: genomics, evolution, and mechanisms. *Cell*, **154**, 26-
687 46.
688 3. Guo, C.J., Xu, G. and Chen, L.L. (2020) Mechanisms of Long Noncoding RNA Nuclear Retention.
689 *Trends Biochem Sci*, **45**, 947-960.
690 4. Zuckerman, B. and Ulitsky, I. (2019) Predictive models of subcellular localization of long RNAs.
691 *RNA*, **25**, 557-572.
692 5. Clark, M.B., Johnston, R.L., Inostroza-Ponta, M., Fox, A.H., Fortini, E., Moscato, P., Dinger, M.E.
693 and Mattick, J.S. (2012) Genome-wide analysis of long noncoding RNA stability. *Genome Res*, **22**,
694 885-898.
695 6. Cabili, M.N., Dunagin, M.C., McClanahan, P.D., Biaesch, A., Padovan-Merhar, O., Regev, A., Rinn,
696 J.L. and Raj, A. (2015) Localization and abundance analysis of human lncRNAs at single-cell and
697 single-molecule resolution. *Genome Biol*, **16**, 20.
698 7. Tilgner, H., Knowles, D.G., Johnson, R., Davis, C.A., Chakrabortty, S., Djebali, S., Curado, J.,
699 Snyder, M., Gingeras, T.R. and Guigo, R. (2012) Deep sequencing of subcellular RNA fractions
700 shows splicing to be predominantly co-transcriptional in the human genome but inefficient for
701 lncRNAs. *Genome Res*, **22**, 1616-1625.
702 8. Mele, M., Mattioli, K., Mallard, W., Shechner, D.M., Gerhardinger, C. and Rinn, J.L. (2017)
703 Chromatin environment, transcriptional regulation, and splicing distinguish lincRNAs and
704 mRNAs. *Genome Res*, **27**, 27-37.
705 9. Kumar, M. and Grammatikakis, I. (2025) Causes, Consequences and Challenges of Intron
706 Retention in lncRNAs. *Mol Cell Biol*, **45**, 576-584.
707 10. Mukherjee, N., Calviello, L., Hirsekorn, A., de Pretis, S., Pelizzola, M. and Ohler, U. (2017)
708 Integrative classification of human coding and noncoding genes through RNA metabolism
709 profiles. *Nat Struct Mol Biol*, **24**, 86-96.
710 11. Hartford, C.C.R., Shrestha, R.L., Pongor, L., Zhao, Y., Chen, X., Fromont, C., Chaudhary, R., Li, X.L.,
711 Pasterczyk, K.R., Kumar, R. *et al.* (2022) Context-Dependent Function of Long Noncoding RNA
712 PURPL in Transcriptome Regulation during p53 Activation. *Mol Cell Biol*, **42**, e0028922.
713 12. Black, D.L. (2003) Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem*,
714 **72**, 291-336.
715 13. Wilkinson, M.E., Charenton, C. and Nagai, K. (2020) RNA Splicing by the Spliceosome. *Annu Rev*
716 *Biochem*, **89**, 359-388.
717 14. Ruskin, B., Zamore, P.D. and Green, M.R. (1988) A factor, U2AF, is required for U2 snRNP binding
718 and splicing complex assembly. *Cell*, **52**, 207-219.
719 15. Berglund, J.A., Chua, K., Abovich, N., Reed, R. and Rosbash, M. (1997) The splicing factor BBP
720 interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. *Cell*, **89**, 781-787.
721 16. Berglund, J.A., Abovich, N. and Rosbash, M. (1998) A cooperative interaction between U2AF65
722 and mBBP/SF1 facilitates branchpoint region recognition. *Genes Dev*, **12**, 858-867.
723 17. Wu, S., Romfo, C.M., Nilsen, T.W. and Green, M.R. (1999) Functional recognition of the 3' splice
724 site AG by the splicing factor U2AF35. *Nature*, **402**, 832-835.
725 18. Zorio, D.A. and Blumenthal, T. (1999) Both subunits of U2AF recognize the 3' splice site in
726 Caenorhabditis elegans. *Nature*, **402**, 835-838.

727 19. Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth,
728 G.P. and Burge, C.B. (2008) Alternative isoform regulation in human tissue transcriptomes.
729 *Nature*, **456**, 470-476.
730 20. Nilsen, T.W. and Graveley, B.R. (2010) Expansion of the eukaryotic proteome by alternative
731 splicing. *Nature*, **463**, 457-463.
732 21. Pan, Q., Shai, O., Lee, L.J., Frey, B.J. and Blencowe, B.J. (2008) Deep surveying of alternative
733 splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet*, **40**,
734 1413-1415.
735 22. Gier, R.A., Budinich, K.A., Evitt, N.H., Cao, Z., Freilich, E.S., Chen, Q., Qi, J., Lan, Y., Kohli, R.M. and
736 Shi, J. (2020) High-performance CRISPR-Cas12a genome editing for combinatorial genetic
737 screening. *Nat Commun*, **11**, 3455.
738 23. Braunschweig, U., Barbosa-Morais, N.L., Pan, Q., Nachman, E.N., Alipanahi, B., Gonatopoulos-
739 Pournatzis, T., Frey, B., Irimia, M. and Blencowe, B.J. (2014) Widespread intron retention in
740 mammals functionally tunes transcriptomes. *Genome Res*, **24**, 1774-1786.
741 24. Wong, J.J., Ritchie, W., Ebner, O.A., Selbach, M., Wong, J.W., Huang, Y., Gao, D., Pinello, N.,
742 Gonzalez, M., Baidya, K. *et al.* (2013) Orchestrated intron retention regulates normal
743 granulocyte differentiation. *Cell*, **154**, 583-595.
744 25. Monteuuis, G., Wong, J.J.L., Bailey, C.G., Schmitz, U. and Rasko, J.E.J. (2019) The changing
745 paradigm of intron retention: regulation, ramifications and recipes. *Nucleic Acids Res*, **47**, 11497-
746 11513.
747 26. Eom, T., Zhang, C., Wang, H., Lay, K., Fak, J., Noebels, J.L. and Darnell, R.B. (2013) NOVA-
748 dependent regulation of cryptic NMD exons controls synaptic protein levels after seizure. *Elife*,
749 **2**, e00178.
750 27. Hossain, M.A., Claggett, J.M., Edwards, S.R., Shi, A., Pennebaker, S.L., Cheng, M.Y., Hasty, J. and
751 Johnson, T.L. (2016) Posttranscriptional Regulation of Gcr1 Expression and Activity Is Crucial for
752 Metabolic Adjustment in Response to Glucose Availability. *Mol Cell*, **62**, 346-358.
753 28. Cha, S., Hong, C.P., Kang, H.A. and Hahn, J.S. (2021) Differential activation mechanisms of two
754 isoforms of Gcr1 transcription factor generated from spliced and un-spliced transcripts in
755 Saccharomyces cerevisiae. *Nucleic Acids Res*, **49**, 745-759.
756 29. Schmitz, U., Pinello, N., Jia, F., Alasmari, S., Ritchie, W., Keightley, M.C., Shini, S., Lieschke, G.J.,
757 Wong, J.J. and Rasko, J.E.J. (2017) Intron retention enhances gene regulatory complexity in
758 vertebrates. *Genome Biol*, **18**, 216.
759 30. Dvinge, H. and Bradley, R.K. (2015) Widespread intron retention diversifies most cancer
760 transcriptomes. *Genome Med*, **7**, 45.
761 31. Boutz, P.L., Bhutkar, A. and Sharp, P.A. (2015) Detained introns are a novel, widespread class of
762 post-transcriptionally spliced introns. *Genes Dev*, **29**, 63-80.
763 32. Gutschner, T., Hammerle, M., Eissmann, M., Hsu, J., Kim, Y., Hung, G., Revenko, A., Arun, G.,
764 Stentrup, M., Gross, M. *et al.* (2013) The noncoding RNA MALAT1 is a critical regulator of the
765 metastasis phenotype of lung cancer cells. *Cancer Res*, **73**, 1180-1189.
766 33. Hutchinson, J.N., Ensminger, A.W., Clemson, C.M., Lynch, C.R., Lawrence, J.B. and Chess, A.
767 (2007) A screen for nuclear transcripts identifies two linked noncoding RNAs associated with
768 SC35 splicing domains. *BMC Genomics*, **8**, 39.
769 34. Ji, P., Diederichs, S., Wang, W., Boing, S., Metzger, R., Schneider, P.M., Tidow, N., Brandt, B.,
770 Buerger, H., Bulk, E. *et al.* (2003) MALAT-1, a novel noncoding RNA, and thymosin beta4 predict
771 metastasis and survival in early-stage non-small cell lung cancer. *Oncogene*, **22**, 8031-8041.
772 35. Bernard, D., Prasanth, K.V., Tripathi, V., Colasse, S., Nakamura, T., Xuan, Z., Zhang, M.Q., Sedel,
773 F., Jourdren, L., Coulpier, F. *et al.* (2010) A long nuclear-retained non-coding RNA regulates
774 synaptogenesis by modulating gene expression. *EMBO J*, **29**, 3082-3093.

775 36. Tripathi, V., Ellis, J.D., Shen, Z., Song, D.Y., Pan, Q., Watt, A.T., Freier, S.M., Bennett, C.F., Sharma,
776 A., Bubulya, P.A. *et al.* (2010) The nuclear-retained noncoding RNA MALAT1 regulates
777 alternative splicing by modulating SR splicing factor phosphorylation. *Mol Cell*, **39**, 925-938.
778 37. Shinn, M.K., Tomares, D.T., Liu, V., Pant, A., Qiu, Y., Vitalis, A., Song, Y.J., Ayala, Y., Ruff, K.M.,
779 Strout, G.W. *et al.* (2026) Nuclear speckle proteins form intrinsic and MALAT1-dependent
780 microphases. *Cell*, **189**, 832-852 e824.
781 38. Spector, D.L. and Lamond, A.I. (2011) Nuclear speckles. *Cold Spring Harb Perspect Biol*, **3**.
782 39. Sharma, A., Takata, H., Shibahara, K., Bubulya, A. and Bubulya, P.A. (2010) Son is essential for
783 nuclear speckle organization and cell cycle progression. *Mol Biol Cell*, **21**, 650-663.
784 40. Nakagawa, S., Ip, J.Y., Shioi, G., Tripathi, V., Zong, X., Hirose, T. and Prasanth, K.V. (2012) Malat1
785 is not an essential component of nuclear speckles in mice. *RNA*, **18**, 1487-1499.
786 41. Zhang, B., Arun, G., Mao, Y.S., Lazar, Z., Hung, G., Bhattacharjee, G., Xiao, X., Booth, C.J., Wu, J.,
787 Zhang, C. *et al.* (2012) The lncRNA Malat1 is dispensable for mouse development but its
788 transcription plays a cis-regulatory role in the adult. *Cell Rep*, **2**, 111-123.
789 42. Li, X.L., Subramanian, M., Jones, M.F., Chaudhary, R., Singh, D.K., Zong, X., Gryder, B., Sindri, S.,
790 Mo, M., Schetter, A. *et al.* (2017) Long Noncoding RNA PURPL Suppresses Basal p53 Levels and
791 Promotes Tumorigenicity in Colorectal Cancer. *Cell Rep*, **20**, 2408-2423.
792 43. Behera, A.K., Kim, J.J., Kordale, S., Pekovic, F., Damodaran, A.P., Kumari, B., Vidak, S., Dickson, E.,
793 Xiao, M.S., Duncan, G. *et al.* (2026) RNA-coupled CRISPR screens reveal ZNF207 as a regulator of
794 LMNA aberrant splicing in progeria. *Mol Cell*, **86**, 41-59 e15.
795 44. Aregger, M., Xing, K. and Gonatopoulos-Pournatzis, T. (2021) Application of CHyMErA Cas9-
796 Cas12a combinatorial genome-editing platform for genetic interaction mapping and gene
797 fragment deletion screening. *Nat Protoc*, **16**, 4722-4765.
798 45. Gonatopoulos-Pournatzis, T., Aregger, M., Brown, K.R., Farhangmehr, S., Braunschweig, U.,
799 Ward, H.N., Ha, K.C.H., Weiss, A., Billmann, M., Durbic, T. *et al.* (2020) Genetic interaction
800 mapping and exon-resolution functional genomics with a hybrid Cas9-Cas12a platform. *Nat*
801 *Biotechnol*, **38**, 638-648.
802 46. Xiao, M.S., Damodaran, A.P., Kumari, B., Dickson, E., Xing, K., On, T.A., Parab, N., King, H.E.,
803 Perez, A.R., Guiblet, W.M. *et al.* (2024) Genome-scale exon perturbation screens uncover exons
804 critical for cell fitness. *Mol Cell*, **84**, 2553-2572 e2519.
805 47. Van Nostrand, E.L., Pratt, G.A., Shishkin, A.A., Gelboin-Burkhart, C., Fang, M.Y., Sundararaman,
806 B., Blue, S.M., Nguyen, T.B., Surka, C., Elkins, K. *et al.* (2016) Robust transcriptome-wide
807 discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat Methods*, **13**,
808 508-514.
809 48. Consortium, E.P. (2012) An integrated encyclopedia of DNA elements in the human genome.
810 *Nature*, **489**, 57-74.
811 49. Luo, Y., Hitz, B.C., Gabdank, I., Hilton, J.A., Kagda, M.S., Lam, B., Myers, Z., Sud, P., Jou, J., Lin, K.
812 *et al.* (2020) New developments on the Encyclopedia of DNA Elements (ENCODE) data portal.
813 *Nucleic Acids Res*, **48**, D882-D889.
814 50. Berhane, T., Holm, A., Karstensen, K.T., Petri, A., Ilieva, M.S., Krarup, H., Vyberg, M., Lovendorf,
815 M.B. and Kauppinen, S. (2022) Knockdown of the long noncoding RNA PURPL induces apoptosis
816 and sensitizes liver cancer cells to doxorubicin. *Sci Rep*, **12**, 19502.
817 51. Han, S., Li, X., Wang, K., Zhu, D., Meng, B., Liu, J., Liang, X., Jin, Y., Liu, X., Wen, Q. *et al.* (2021)
818 PURPL represses autophagic cell death to promote cutaneous melanoma by modulating ULK1
819 phosphorylation. *Cell Death Dis*, **12**, 1070.
820 52. Middleton, R., Gao, D., Thomas, A., Singh, B., Au, A., Wong, J.J., Bomane, A., Cosson, B., Eyras, E.,
821 Rasko, J.E. *et al.* (2017) IRFinder: assessing the impact of intron retention on mammalian gene
822 expression. *Genome Biol*, **18**, 51.

823 53. Martinez-Terroba, E., Plasek-Hegde, L.M., Chiotakakos, I., Li, V., de Miguel, F.J., Robles-Oteiza,
824 C., Tyagi, A., Politi, K., Zamudio, J.R. and Dimitrova, N. (2024) Overexpression of Malat1 drives
825 metastasis through inflammatory reprogramming of the tumor microenvironment. *Sci Immunol*,
826 **9**, eadh5462.
827 54. Arun, G., Aggarwal, D. and Spector, D.L. (2020) MALAT1 Long Non-Coding RNA: Functional
828 Implications. *Noncoding RNA*, **6**.
829 55. Jacob, A.G. and Smith, C.W.J. (2017) Intron retention as a component of regulated gene
830 expression programs. *Hum Genet*, **136**, 1043-1057.
831 56. Shao, C., Yang, B., Wu, T., Huang, J., Tang, P., Zhou, Y., Zhou, J., Qiu, J., Jiang, L., Li, H. *et al.*
832 (2014) Mechanisms for U2AF to define 3' splice sites and regulate alternative splicing in the
833 human genome. *Nat Struct Mol Biol*, **21**, 997-1005.
834 57. Naftelberg, S., Schor, I.E., Ast, G. and Kornblihtt, A.R. (2015) Regulation of alternative splicing
835 through coupling with transcription and chromatin structure. *Annu Rev Biochem*, **84**, 165-198.
836 58. Herzel, L., Ottoz, D.S.M., Alpert, T. and Neugebauer, K.M. (2017) Splicing and transcription touch
837 base: co-transcriptional spliceosome assembly and function. *Nat Rev Mol Cell Biol*, **18**, 637-650.
838 59. Wong, J.J., Gao, D., Nguyen, T.V., Kwok, C.T., van Geldermalsen, M., Middleton, R., Pinello, N.,
839 Thoeng, A., Nagarajah, R., Holst, J. *et al.* (2017) Intron retention is regulated by altered MeCP2-
840 mediated splicing factor recruitment. *Nat Commun*, **8**, 15134.
841 60. Guo, R., Zheng, L., Park, J.W., Lv, R., Chen, H., Jiao, F., Xu, W., Mu, S., Wen, H., Qiu, J. *et al.* (2014)
842 BS69/ZMYND11 reads and connects histone H3.3 lysine 36 trimethylation-decorated chromatin
843 to regulated pre-mRNA processing. *Mol Cell*, **56**, 298-310.
844 61. Krchnakova, Z., Thakur, P.K., Krausova, M., Bieberstein, N., Haberman, N., Muller-McNicoll, M.
845 and Stanek, D. (2019) Splicing of long non-coding RNAs primarily depends on polypyrimidine
846 tract and 5' splice-site sequences due to weak interactions with SR proteins. *Nucleic Acids Res*,
847 **47**, 911-928.
848 62. Dumbovic, G., Braunschweig, U., Langner, H.K., Smallegan, M., Biayna, J., Hass, E.P., Jastrzebska,
849 K., Blencowe, B., Cech, T.R., Caruthers, M.H. *et al.* (2021) Nuclear compartmentalization of TERT
850 mRNA and TUG1 lncRNA is driven by intron retention. *Nat Commun*, **12**, 3308.
851 63. Desideri, F., Cipriano, A., Petrezselyova, S., Buonaiuto, G., Santini, T., Kasparek, P., Prochazka, J.,
852 Janson, G., Paiardini, A., Calicchio, A. *et al.* (2020) Intronic Determinants Coordinate Charme
853 lncRNA Nuclear Activity through the Interaction with MATR3 and PTBP1. *Cell Rep*, **33**, 108548.
854 64. Guo, C.J., Ma, X.K., Xing, Y.H., Zheng, C.C., Xu, Y.F., Shan, L., Zhang, J., Wang, S., Wang, Y.,
855 Carmichael, G.G. *et al.* (2020) Distinct Processing of lncRNAs Contributes to Non-conserved
856 Functions in Stem Cells. *Cell*, **181**, 621-636 e622.
857 65. Barutcu, A.R., Wu, M., Braunschweig, U., Dyakov, B.J.A., Luo, Z., Turner, K.M., Durbic, T., Lin,
858 Z.Y., Weatheritt, R.J., Maass, P.G. *et al.* (2022) Systematic mapping of nuclear domain-associated
859 transcripts reveals speckles and lamina as hubs of functionally distinct retained introns. *Mol Cell*,
860 **82**, 1035-1052 e1039.
861 66. Kolberg, L., Raudvere, U., Kuzmin, I., Adler, P., Vilo, J. and Peterson, H. (2023) g:Profiler-
862 interoperable web service for functional enrichment analysis and gene identifier mapping (2023
863 update). *Nucleic Acids Res*, **51**, W207-W212.
864 67. Lee, J.S., Dan, T., Zhang, H., Cheng, Y., Rehfeld, F., Brugarolas, J. and Mendell, J.T. (2025) An
865 ultraconserved snoRNA-like element in long noncoding RNA CRNDE promotes ribosome
866 biogenesis and cell proliferation. *Mol Cell*, **85**, 1543-1560 e1510.
867 68. Brown, A., Winter, J., Gapinske, M., Tague, N., Woods, W.S. and Perez-Pinera, P. (2019)
868 Multiplexed and tunable transcriptional activation by promoter insertion using nuclease-assisted
869 vector integration. *Nucleic Acids Res*, **47**, e67.

870 69. Brown, A., Woods, W.S. and Perez-Pinera, P. (2016) Multiplexed Targeted Genome Engineering
871 Using a Universal Nuclease-Assisted Vector Integration System. *ACS Synth Biol*, **5**, 582-588.
872 70. Brown, A., Woods, W.S. and Perez-Pinera, P. (2017) Targeted Gene Activation Using RNA-Guided
873 Nucleases. *Methods Mol Biol*, **1468**, 235-250.

874

875 **Figure legends**

876 **Figure 1. CRISPR-based screen identifies U2AF2 as a promoter of intron retention. (A)**
877 Schematic of the *PURPL* IR reporter used in the screen to identify IR regulators. The pLCHKO
878 vector contains the puromycin resistance gene used for selection. The transgene also contains the
879 genomic sequences of exon 2 (161 bp), intron 2 (1,943 bp), and exon 3 (163 bp) of the *PURPL*
880 gene under a Doxycycline-inducible promoter. A hybrid guide (hg)RNA for guiding *Sp*Cas9 and
881 *Lb*Cas12a is expressed under a U6 promoter. Top arrow: The hgRNA is cleaved by Cas12a to
882 produce 2 guide RNAs for dual targeting of each protein coding gene, paralog targeting, and
883 genetic interaction mapping. Bottom arrow: Upon intron splicing or intron retention, two different
884 reads are generated. **(B)** Two different cell lines, stably expressing Cas9 and Cas12a, are used:
885 HAP1, and RPE1. The CRASP-seq pipeline until Illumina paired-end sequencing of CRISPR
886 libraries is depicted. **(C)** Heatmap showing single genes only (after removing gene pairs), whether
887 identified individually or as part of a gene pair. The hits are ordered by the average ΔPIR values
888 across HAP1 and RPE1, from lowest to highest. **(D)** Left: Gene ontology enrichment analysis of
889 biological processes regulated by the hits of the screen. The most significant pathways involve
890 mRNA processing and RNA splicing. Right: Gene ontology enrichment analysis of cellular
891 components regulated by the hits of the screen. These components include mainly Spliceosomal
892 Complex, small nuclear ribonucleoprotein complex and the Catalytic step 2 Spliceosome. **(E)**
893 Venn diagram showing that there are 5 common hits which cause a reduction in the percentage of

894 intron retention (ΔPIR) of the splicing reporter in the two cell lines. **(F)** Table showing the top 5
895 hits of the screen with a ΔPIR for each hit in each cell line and their full names.

896 **Figure 2. U2AF2 binds to intron 2-containing *PURPL* transcripts and promotes intron 2**
897 **retention. (A)** IGV snapshot of eCLIP-seq data showing binding sites and enrichment of U2AF2
898 and U2AF1 on the *PURPL* transcripts around intron 2. eCLIP-seq data for PTBP1, PRPF8, and
899 SRSF1 are indicated. The annotated locus by RefSeq is also indicated. eCLIP data-seq was
900 downloaded from encodeproject.org. **(B)** RT-qPCR after RNA-IPs using a U2AF2 antibody with
901 primer pairs specifically detecting *PURPL* transcripts as indicated in Figure S1B. U2AF2 binds to
902 transcripts containing intron 2 but not the ones with intron 1, intron 3, or spliced exons 2 and 3.
903 Samples were normalized to IgG-IP. 18S was used as a loading control. **(C)** Validation of hgRNAs
904 using the *PURPL* minigene in HAP1 and RPE1 cells. The gels show RT-PCR products for *PURPL*
905 transcripts with specific primers for the minigene. The hgRNAs used were targeting intergenic
906 region, or U2AF2. The schematics next to the gel indicate the expected products of the intron-
907 retained and spliced isoforms. **(D)** Western blot for U2AF2 showing successful knockdown of
908 U2AF2 protein in SKHEP1 cells. GAPDH was used as a loading control (lower panel). **(E)**
909 Schematic of the PCR primer triplet used to detect intron 2 retention (red) or splicing (purple). The
910 length for each PCR product is indicated. **(F)** Gel with RT-PCR products for *PURPL* upon
911 knockdown of U2AF2 with 2 different siRNAs in SKHEP1 cells. The schematics next to the gel
912 indicate the expected products of the intron-retained and spliced isoforms. Between the two
913 expected PCR products, we observed an extra band corresponding to the inclusion of an alternative
914 exon inside *PURPL* intron 2 as observed in RefSeq, the inclusion of which is not affected by
915 U2AF2. **(G)** Graph with quantitation of the gel bands in **(F)**. Error bars represent standard
916 deviations from 3 **(G)** experiments. ***$p<0.001$.

917 **Figure 3. Expression of Intron 2 Retaining *PURPL* leads to high proliferation. (A)** RNA-FISH
918 images for *PURPL* with intron 2 retention and *MALAT1* in HCT116 cells without treatment or
919 after 24 h of 2 mM of Hydorxyurea (HU) to induce *PURPL* expression. Scale bar is 10 μm. **(B)**
920 RNA stability assays were performed for *PURPL* transcripts by measuring their levels by RT-
921 qPCR following 0, 2, and 4 h of ActD treatment in SKHEP1 cells. *MYC* was used as a positive
922 control for unstable RNA. **(C)** RT-qPCR showing *PURPL* depletion in in SKHEP1 cells using 3
923 different gRNAs (sg2, sg4, sg5) compared to a Non-Targeting Control gRNA. **(D)** RT-qPCR for
924 intron 2-containing PURPL transcript after 48 h of 1 μg/mL doxycycline treatment in comparison
925 to no treatment in SKHEP1 *PURPL*-CRISPRi populations using the 3 gRNAs as in **(C)**. **(E)**
926 Proliferation assay showing the effect of overexpression of intron 2-containing *PURPL* transcript
927 in the proliferation of SKHEP1 cells where the endogenous *PURPL* is knocked down with
928 CRISPRi. Error bars represent standard deviation from 3 populations with different gRNAs. The
929 cells were treated with 1 μg/mL doxycycline to induce intron 2-retained *PURPL* expression and
930 cell proliferation was monitored at 3 and 6 days. **(F)** Gel with RT-PCR products for transcripts
931 using primers as indicated in Fig. 2E upon doxycycline treatment of SKHEP1 CRIPSRi cells. The
932 three repeats represent 3 different clones of cells. The last lane is an RT- control. The schematics
933 next to the gel indicate the expected products of the intron-retained and spliced isoforms. NS
934 indicates a non-specific band. Error bars in **(B)**, **(C)**, and **(D)** represent standard deviations from 2
935 experiments. *p<0.05, **p<0.01, ***p<0.001.

936 **Figure 4. U2AF2 directly promotes Intron Retention of *MALAT1*. (A)** U2AF2 was knocked
937 down in SKHEP1 cells and 72 h later, RNA was extracted and RNA-seq was performed. *Left*:
938 Number of decreased (blue) and increased (red) IR events at various p-values after U2AF2
939 knockdown as analyzed with the IR Finder algorithm. The purple arrow indicates the *PURPL* IR

event and green arrows indicate *MALAT1* IR events *Right*: Pie chart of the numbers of increased and decreased IR events upon U2AF2 knockdown. **(B)** Floating bar plots showing the IR ratio of *PURPL* intron 2 and the IR ratio of intron 1 (middle) and intron 2 (right) of *MALAT1* in siCTRL and siU2AF2#1 samples as analyzed with the IRFinder algorithm. **(C)** and **(E)** RT-PCR for *MALAT1* using a primer pair flanking the regulated intron 1 **(C)** or intron 2 **(E)** upon knockdown of U2AF2 with 2 different siRNAs in HCT116 and SKHEP1 cells. The schematics next to the gel indicate the expected products of the intron-retained and spliced isoforms. **(D)** and **(F)** Bar graph with quantitation of the gel bands from **(C)** and **(E)** in SKHEP1 cells. Error bars represent standard deviations from 2 independent experiments. *$p<0.05$, **$p<0.01$, ***$p<0.001$.

Figure 5. U2AF2 promotes localization of *MALAT1* to nuclear speckles. (A) RNA-FISH images for *MALAT1* and Immunofluorescence images for SON is shown upon transfection of SKHEP1 cells with siCTRL or siU2AF2 with two different siRNAs. *MALAT1* is enriched in nuclear speckles in the siCTRL but not upon U2AF2 knockdown. **(B)** Quantitation of the speckle to nuclear plasma *MALAT1* signal ratio in the three replicates in panel **(A)**. ****$p<0.0001$.

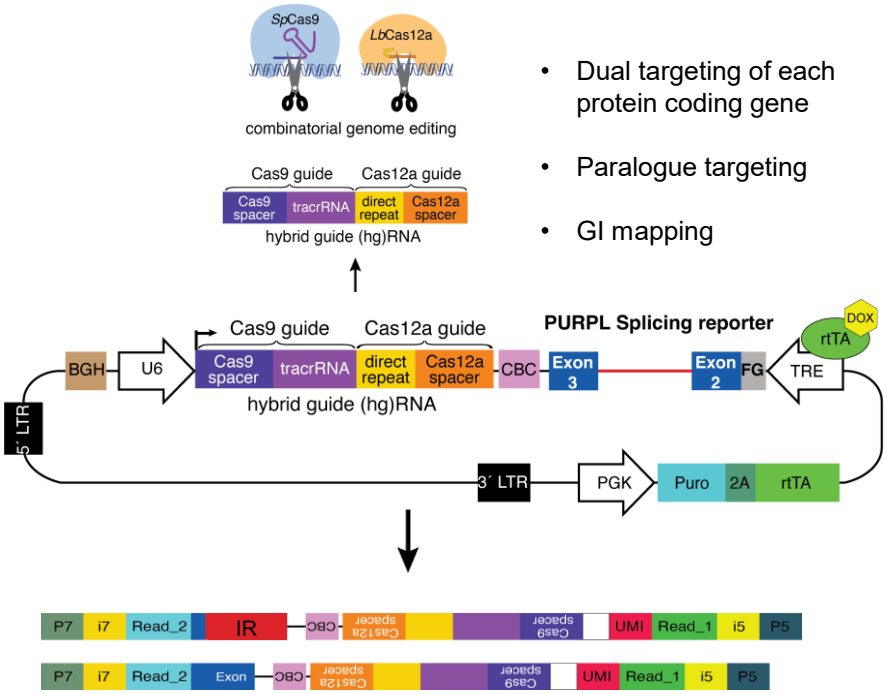
Figure 6. Intron 2 of *MALAT1* dictates its localization to nuclear speckles. (A) RNA-FISH images for *MALAT1* and Immunofluorescence images for SON in a HCT116 *MALAT1*-KO clone (clone 1) where Empty Vector (EV) or constructs that exogenously express *MALAT1* transcripts were re-introduced using a doxy-inducible lentivirus. Expressed *MALAT1* transcripts were full-length (WT) or intron 1 deleted (del-1) or intron 2 deleted (del-2) or both introns deleted simultaneously (del-1+2). The cells were first treated with 1 μg/mL doxycycline for 48 h to induce *MALAT1* expression. *MALAT1* without intron 2 does not enrich in speckles as it is shown with SON protein. Scale bar is 10 μm. **(B)** Graph showing quantification of the percentage of cells

962 where *MALAT1* transcripts are enriched in speckles or being diffused in HCT116 cells clone 1.
963 Error bars represent standard deviations from 2 independent experiments.

964 **Figure 7. Intron 2 regulates the migration capacity of *MALAT1* in breast cancer cells. (A)**
965 RNA-FISH images for *MALAT1* and Immunofluorescence images for SON are shown upon
966 transfection of MDA-MB-231 cells with siCTRL or siU2AF2. *MALAT1* is enriched in nuclear
967 speckles in the siCTRL but not upon U2AF2 knockdown. **(B)** Quantitation of the speckle to nuclear
968 plasma *MALAT1* signal ratio in the three replicates in panel **(A)**. **(C)** RNA-FISH images for
969 *MALAT1* and Immunofluorescence images for SON in MDA-MB-231 clones either WT, or whole
970 locus *MALAT1* deletion, or intron 2 deletion. When intron 2 is deleted, *MALAT1* does not enrich
971 in speckles as it is shown with SON protein. Scale bar is 10 μm. **(D)** Quantitation of the speckle to
972 nuclear plasma *MALAT1* signal ratio in the three replicates in panel **(C)**. **(E)** Graph showing
973 quantification of the percentage of *MALAT1* transcripts enriched in speckles or being diffused in
974 MDA-MB-231 cells in panel **(C)**, N=2. **(F)** Bar graph showing the number of cells that have
975 migrated in transwell migration assays conducted in MDA-MB-231 WT, *MALAT1* KO and
976 *MALAT1* – intron 2 deletion clones. Intron 2 deletion leads to decreased migration potential. The
977 number of cells are the sum of cells from 5 different fields, N=4. *$p<0.05$, **$p<0.01$, ***$p<0.001$,
978 ****$p<0.0001$.

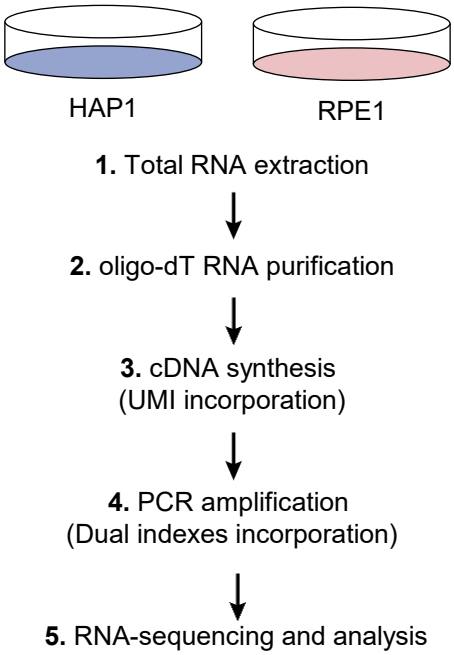
Figure 1

A

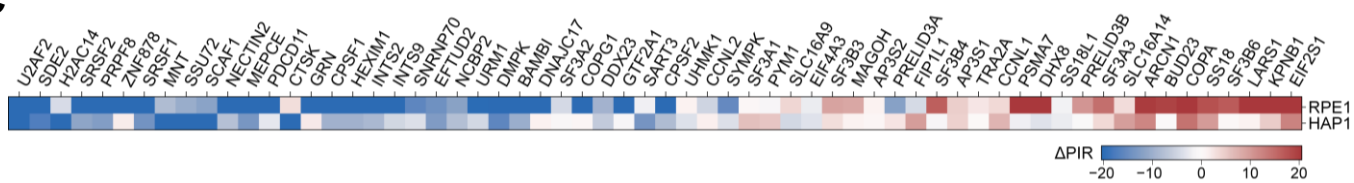


- Dual targeting of each protein coding gene
- Paralogue targeting
- GI mapping

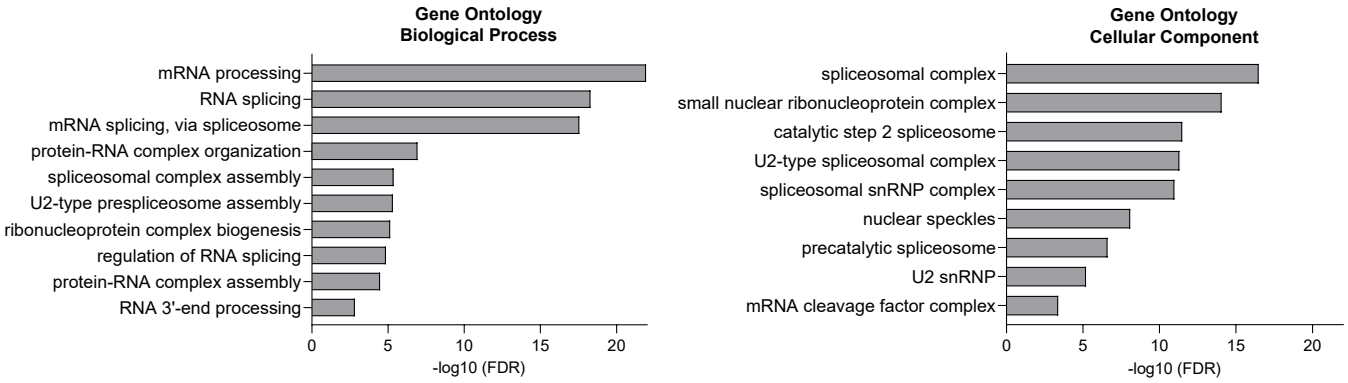
B



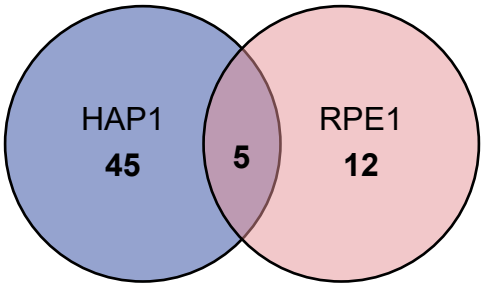
C



D



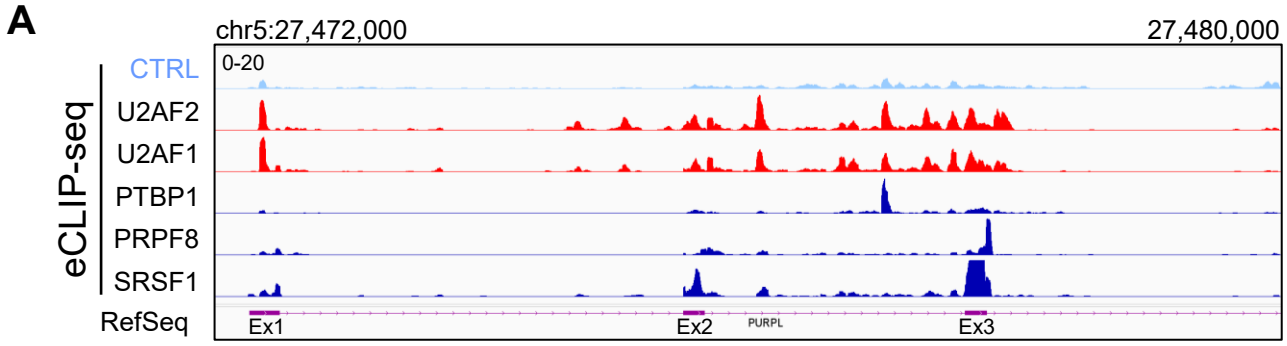
E

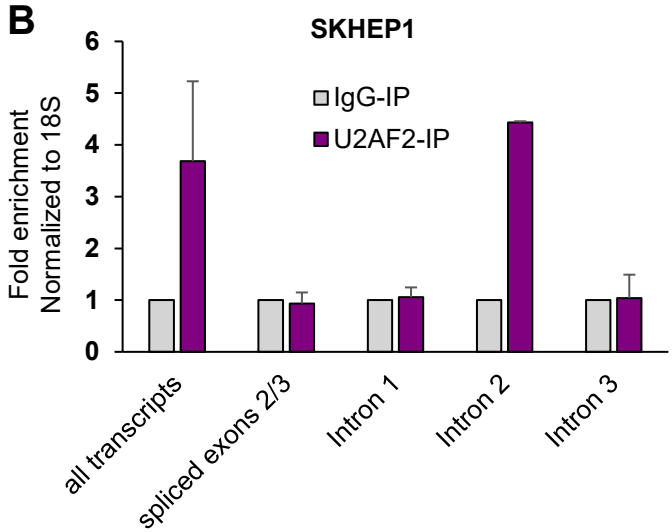


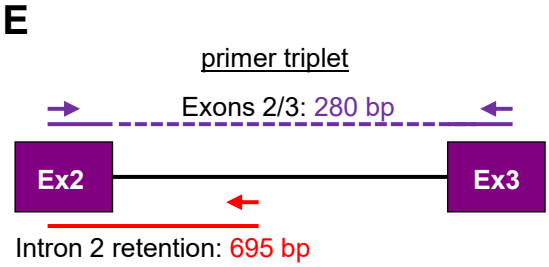
F

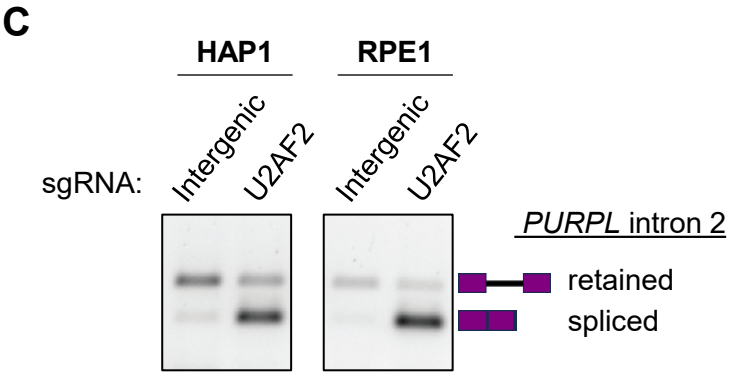
HIT	ΔPIR HAP1	ΔPIR RPE1	Name
U2AF2	-28.42	-23.98	U2 Small Nuclear RNA Auxiliary Factor 2
SNRNP70	-16.05	-3.69	Small Nuclear Ribonucleoprotein U1 Subunit 70
EFTUD2	-14.64	-12.84	Elongation Factor Tu GTP Binding Domain Containing 2
DDX23	-9.49	-6.28	DEAD-Box Helicase 23
SART3	-2.01	-14.01	Spliceosome Associated Factor 3, U4/U6 Recycling Protein

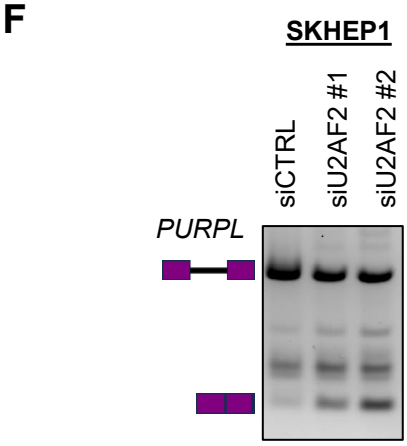
Figure 2

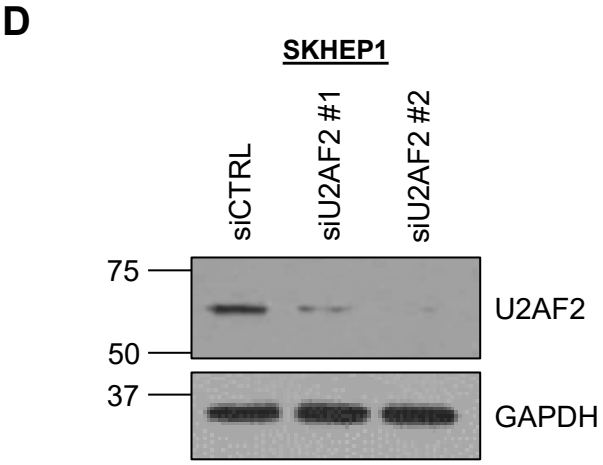

A
chr5:27,472,000
27,480,000
0-20
CTRL
U2AF2
U2AF1
PTBP1
PRPF8
SRSF1
eCLIP-seq
RefSeq
Ex1
Ex2
PURPL
Ex3


B
SKHEP1
IgG-IP
U2AF2-IP
Fold enrichment
Normalized to 18S
0
1
2
3
4
5
6
all transcripts
spliced exons 2/3
Intron 1
Intron 2
Intron 3


E
primer triplet
Exons 2/3: 280 bp
Ex2
Ex3
Intron 2 retention: 695 bp


C
HAP1
RPE1
sgRNA:
Intergenic
U2AF2
Intergenic
U2AF2
PURPL intron 2
retained
spliced


F
SKHEP1
siCTRL
siU2AF2 #1
siU2AF2 #2
PURPL


D
SKHEP1
siCTRL
siU2AF2 #1
siU2AF2 #2
75
50
37
U2AF2
GAPDH

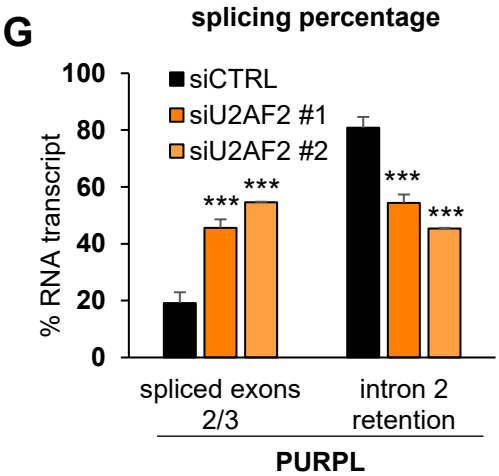
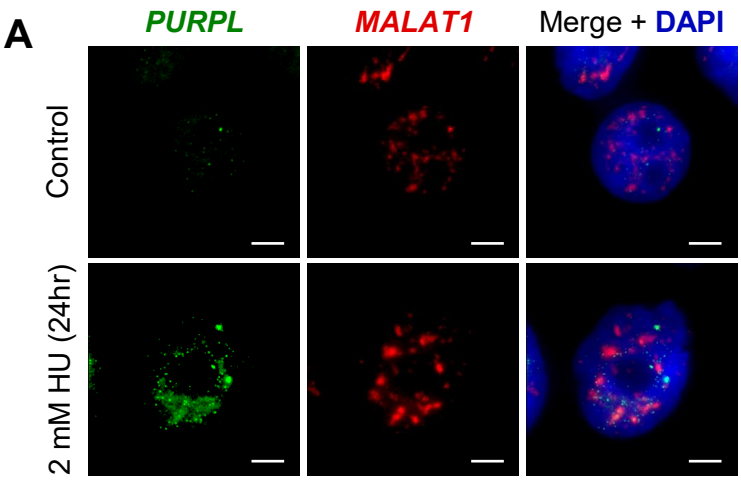
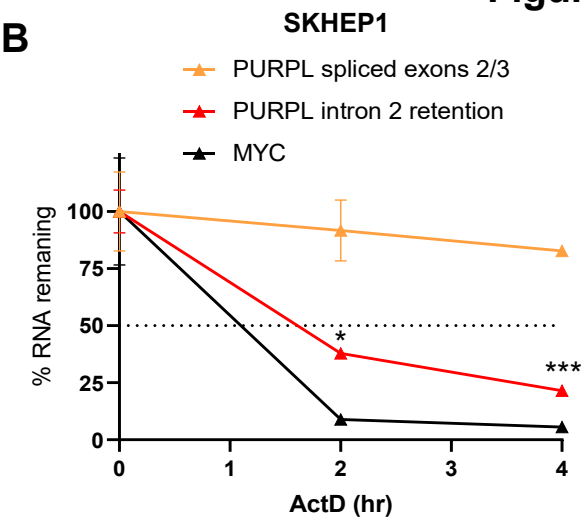
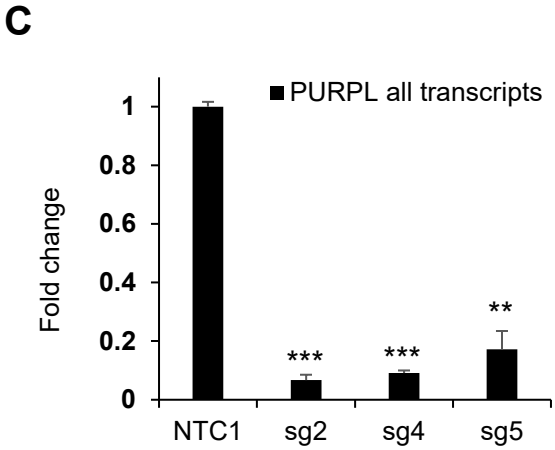
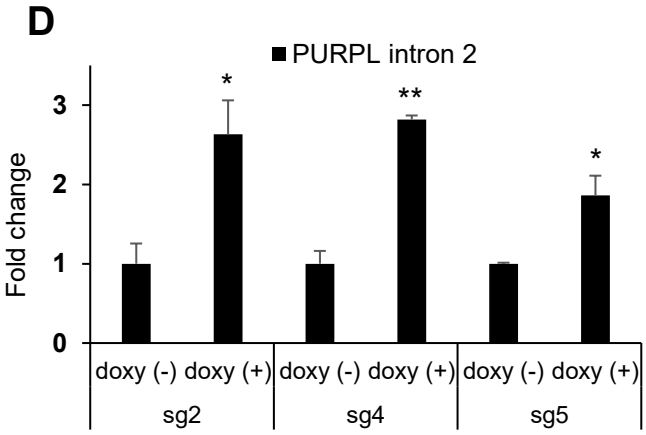

G
splicing percentage
siCTRL
siU2AF2 #1
siU2AF2 #2
% RNA transcript
0
20
40
60
80
100
***
***
***
***
spliced exons 2/3
intron 2 retention
PURPL

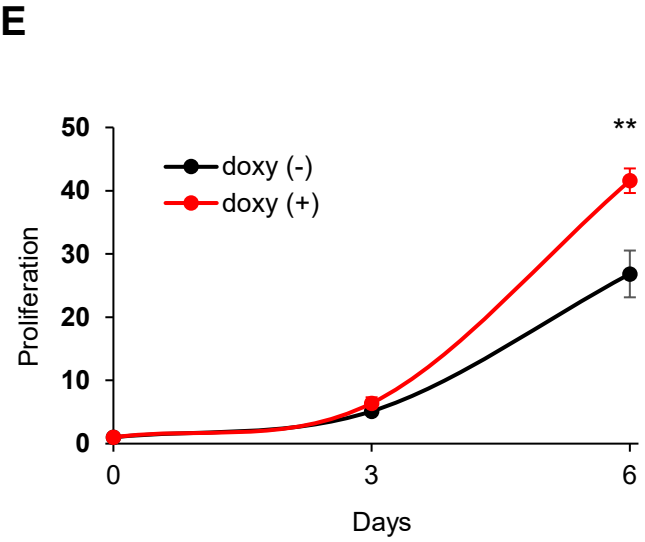
Figure 3


A
PURPL
MALAT1
Merge + DAPI
Control
2 mM HU (24hr)


B
SKHEP1
PURPL spliced exons 2/3
PURPL intron 2 retention
MYC
% RNA remaning
100
75
50
25
0
*
***
0
1
2
3
4
ActD (hr)


C
PURPL all transcripts
Fold change
1
0.8
0.6
0.4
0.2
0
***
***
**
NTC1
sg2
sg4
sg5


D
PURPL intron 2
Fold change
3
2
1
0
*
**
*
doxy (-)
doxy (+)
doxy (-)
doxy (+)
doxy (-)
doxy (+)
sg2
sg4
sg5


E
doxy (-)
doxy (+)
Proliferation
50
40
30
20
10
0
**
0
3
6
Days

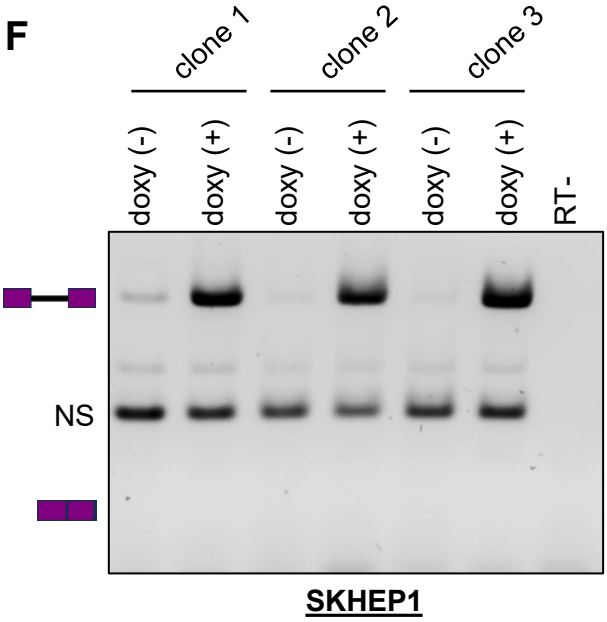
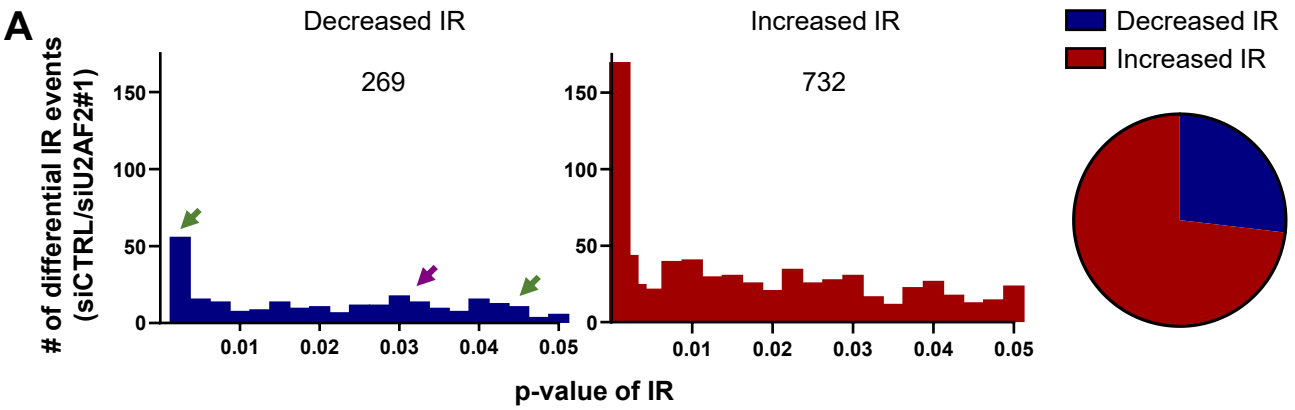
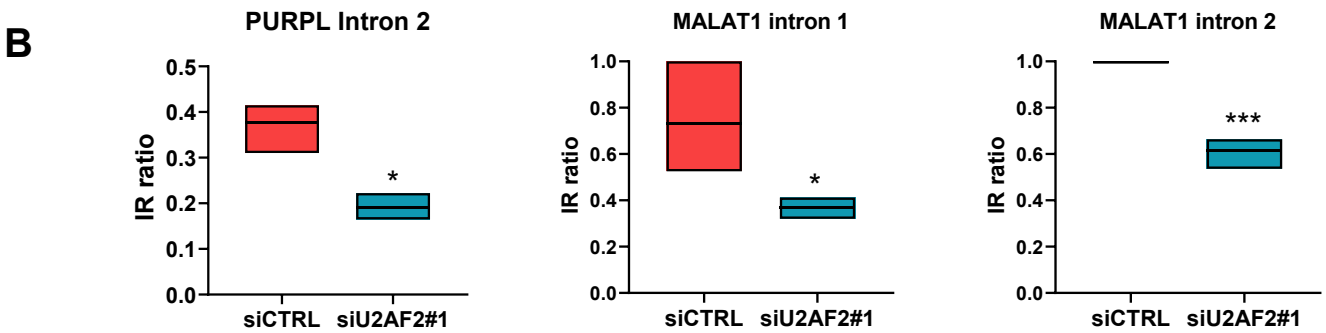
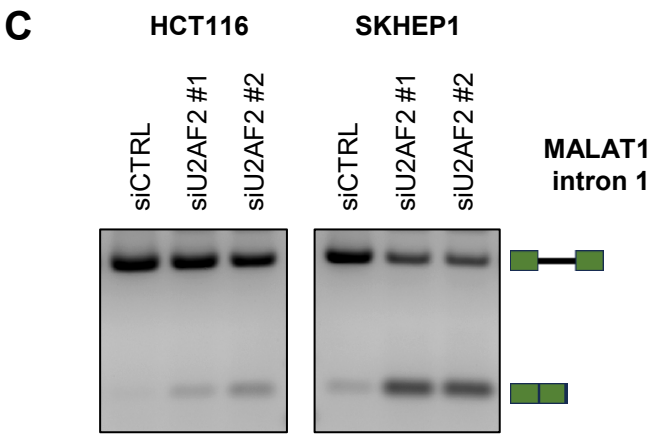
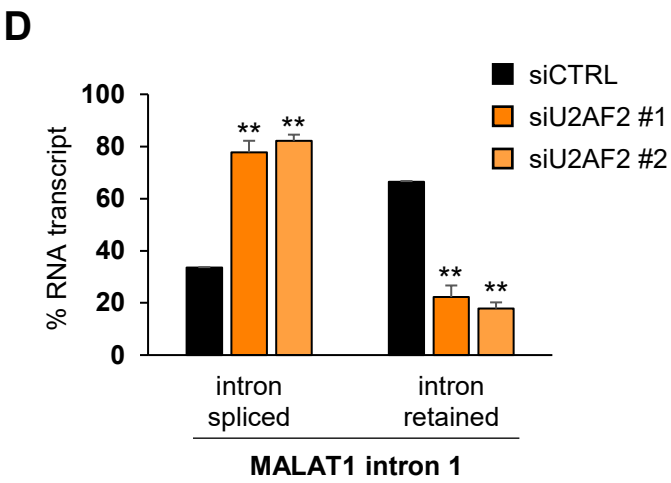

F
clone 1
clone 2
clone 3
doxy (-)
doxy (+)
doxy (-)
doxy (+)
doxy (-)
doxy (+)
RT-
NS
SKHEP1

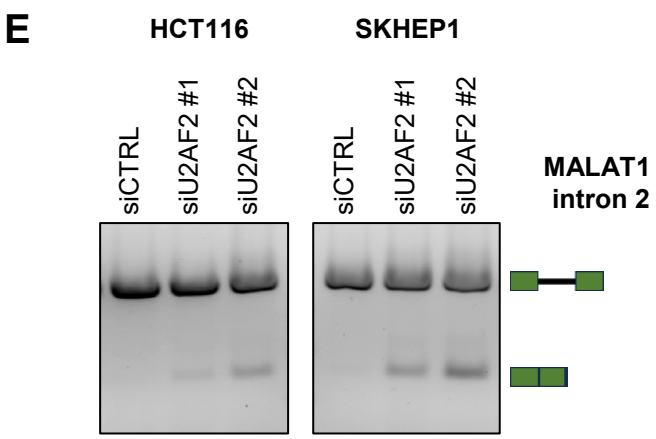
Figure 4


A
Decreased IR
Increased IR
269
732
# of differential IR events (siCTRL/siU2AF2#1)
p-value of IR
Decreased IR
Increased IR


B
PURPL Intron 2
MALAT1 intron 1
MALAT1 intron 2
IR ratio
siCTRL
siU2AF2#1


C
HCT116
SKHEP1
siCTRL
siU2AF2 #1
siU2AF2 #2
MALAT1 intron 1


D
siCTRL
siU2AF2 #1
siU2AF2 #2
% RNA transcript
intron spliced
intron retained
MALAT1 intron 1


E
HCT116
SKHEP1
siCTRL
siU2AF2 #1
siU2AF2 #2
MALAT1 intron 2

F
siCTRL
siU2AF2 #1
siU2AF2 #2
% RNA transcript
intron spliced
intron retained
MALAT1 intron 2

Figure 5

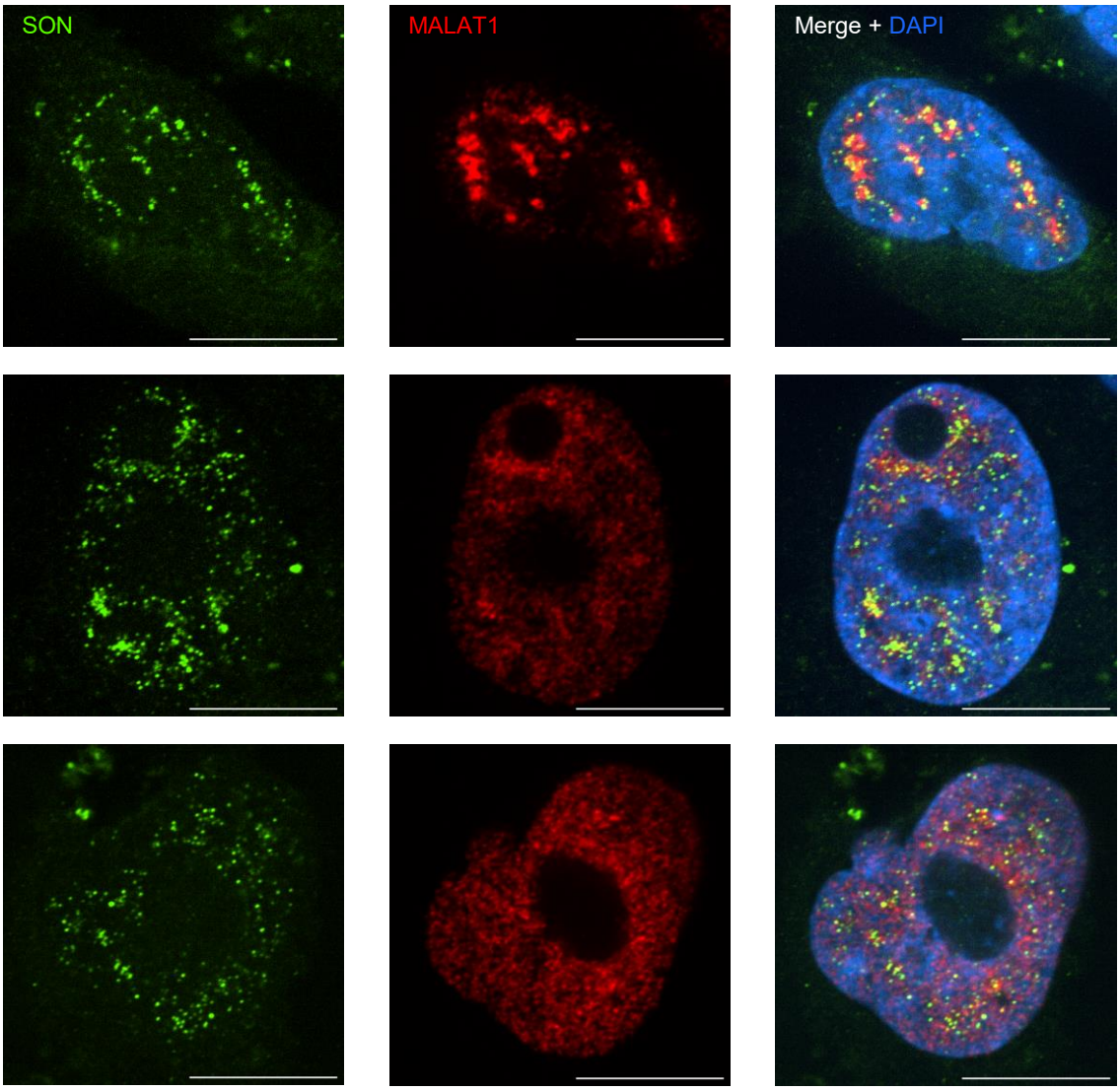
A

SKHEP1

siCTRL

siU2AF2 #1

siU2AF2 #2



B

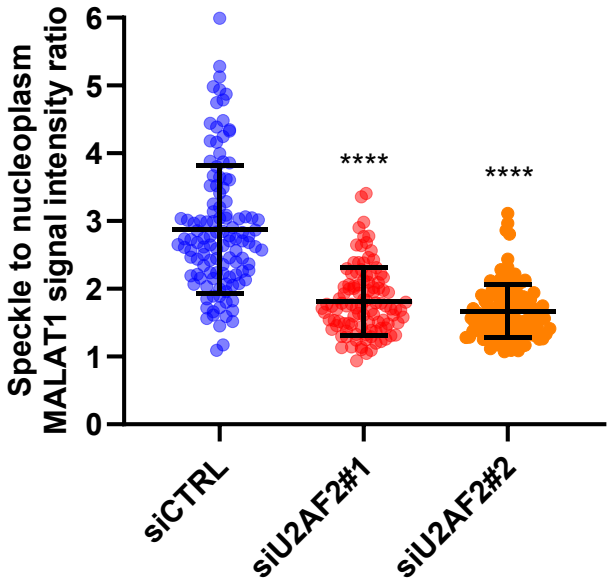
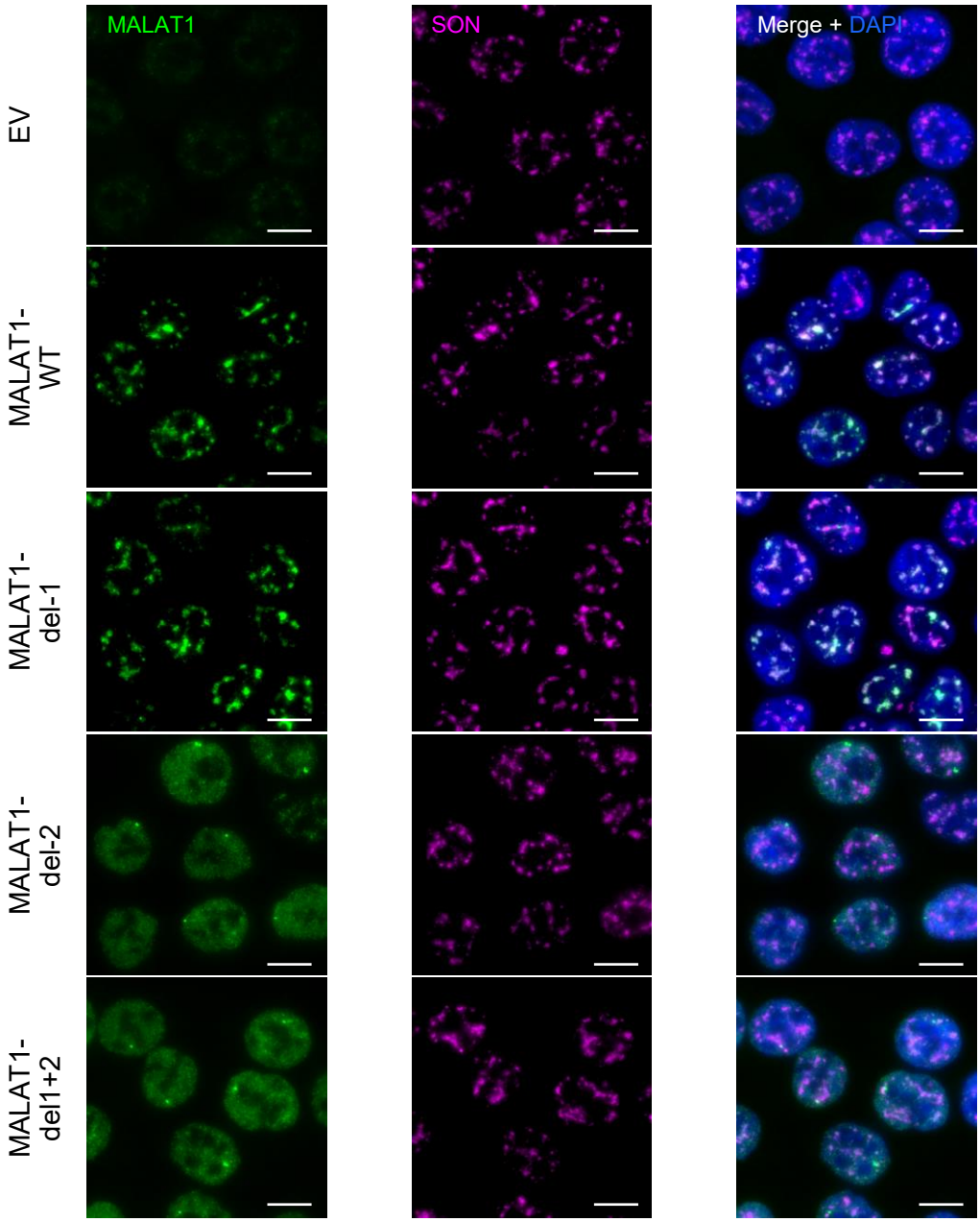


Figure 6

A


MALAT1
SON
Merge + DAPI
EV
MALAT1-WT
MALAT1-del-1
MALAT1-del-2
MALAT1-del1+2

B

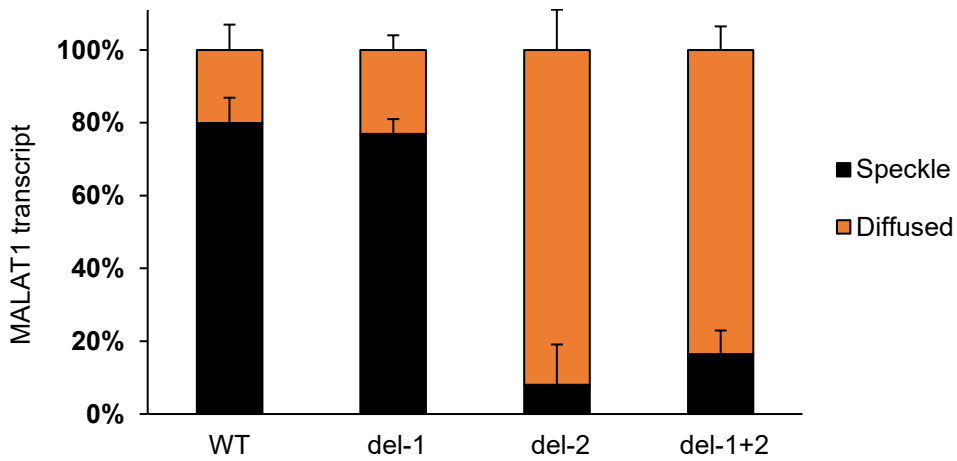
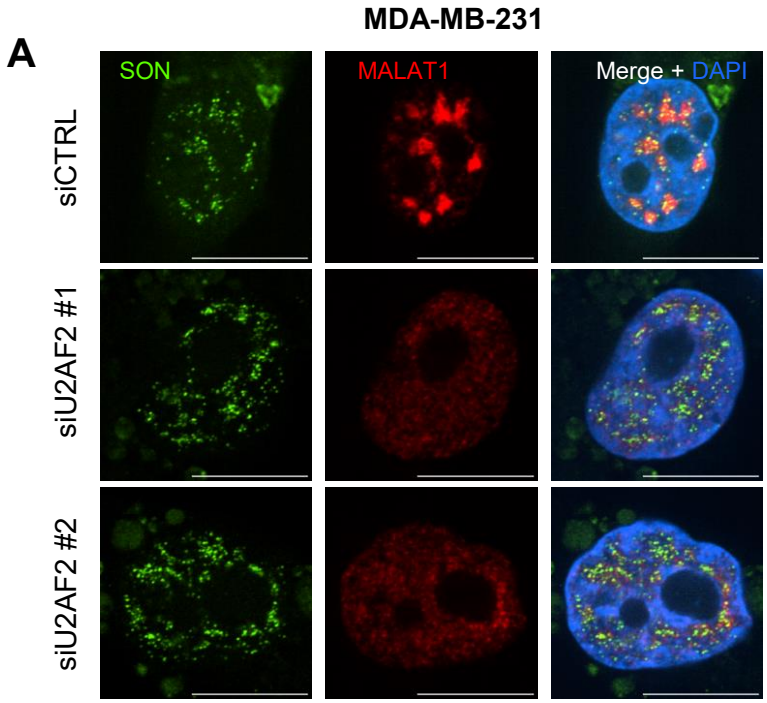
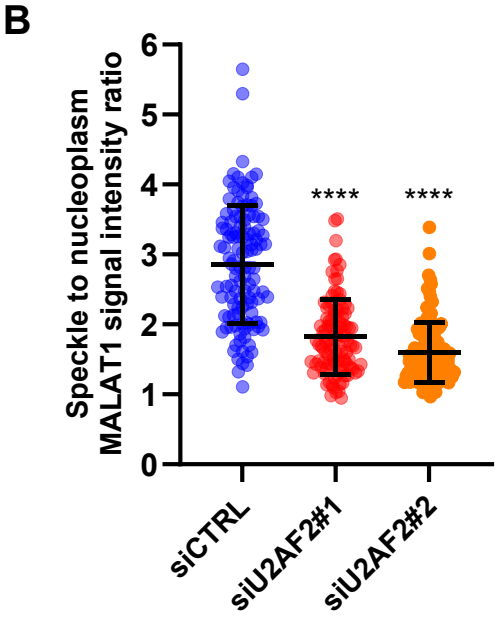
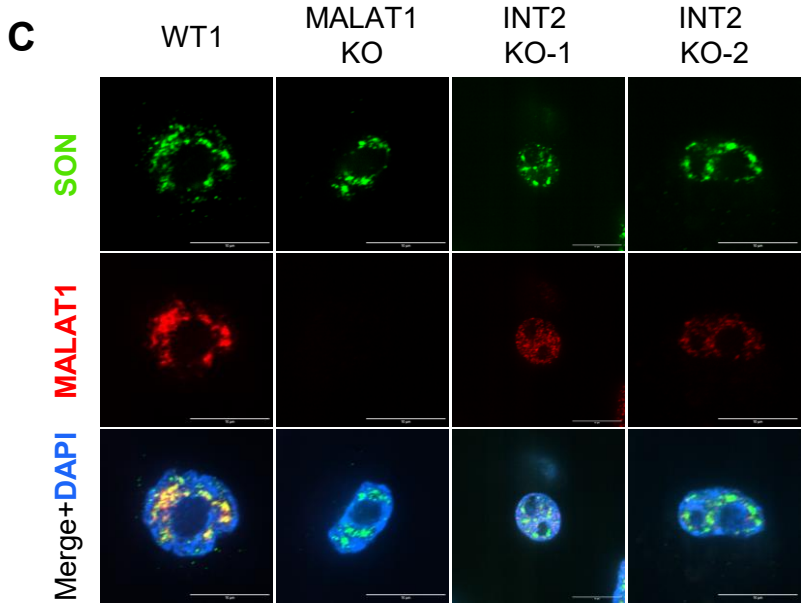
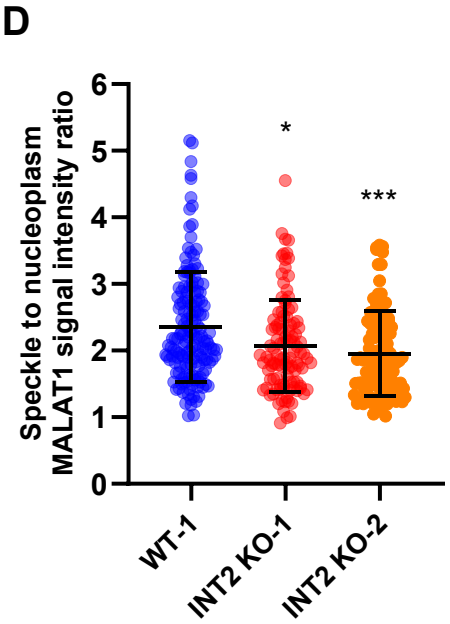

MALAT1 transcript
100%
80%
60%
40%
20%
0%
WT
del-1
del-2
del-1+2
Speckle
Diffused

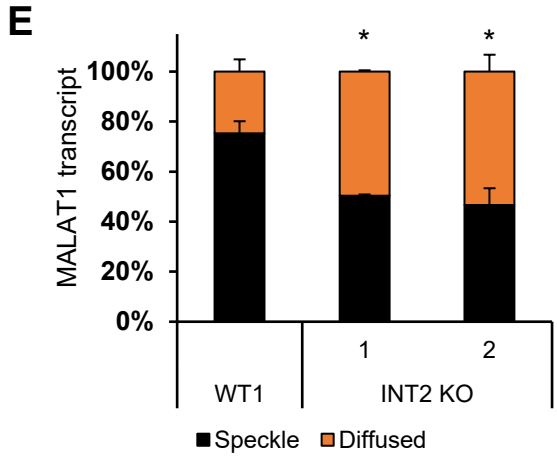
Figure 7

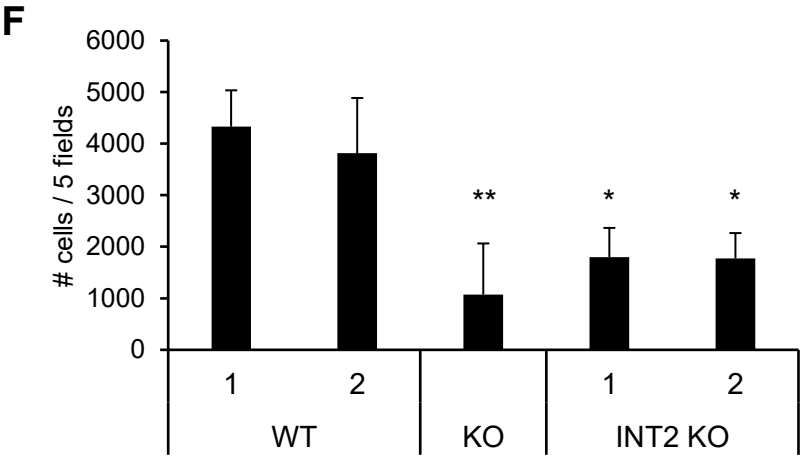

MDA-MB-231
A
SON
MALAT1
Merge + DAPI
siCTRL
siU2AF2 #1
siU2AF2 #2


B
Speckle to nucleoplasm MALAT1 signal intensity ratio
0
1
2
3
4
5
6
****
****
siCTRL
siU2AF2#1
siU2AF2#2


C
WT1
MALAT1 KO
INT2 KO-1
INT2 KO-2
SON
MALAT1
Merge+DAPI


D
Speckle to nucleoplasm MALAT1 signal intensity ratio
0
1
2
3
4
5
6
*
***
WT-1
INT2 KO-1
INT2 KO-2


E
MALAT1 transcript
100%
80%
60%
40%
20%
0%
*
*
1
2
WT1
INT2 KO
Speckle
Diffused


F
cells / 5 fields
6000
5000
4000
3000
2000
1000
0
**
*
*
1
2
1
2
WT
KO
INT2 KO